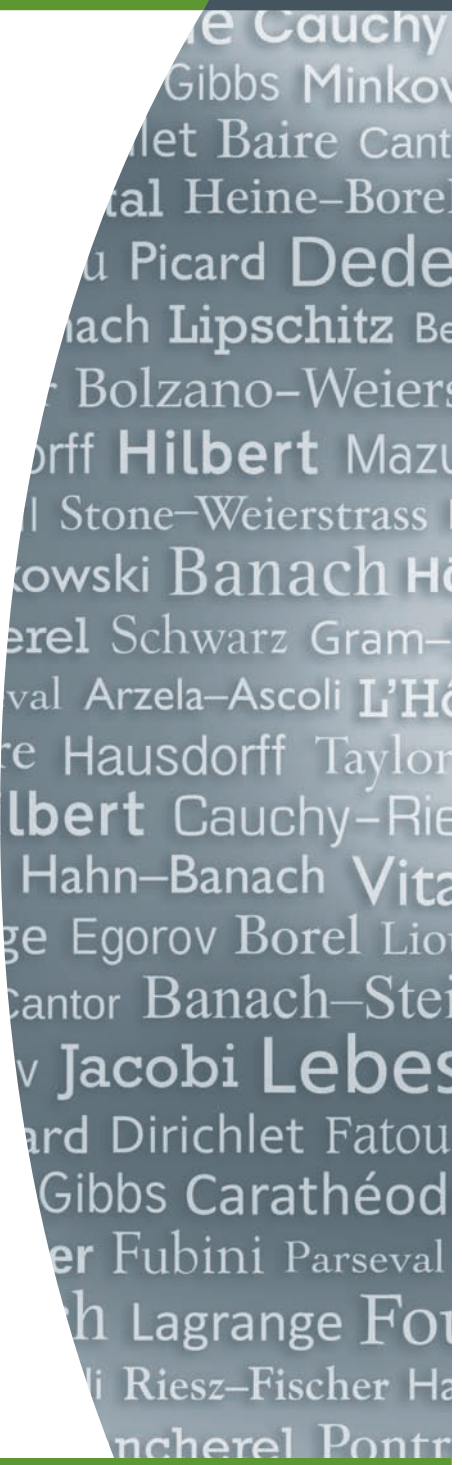
The Sally SERIES

Pure and Applied UNDERGRADUATE TEXTS 20



Fundamentals of Mathematical Analysis

Paul J. Sally, Jr.

American Mathematical Society

Pure and Applied
UNDERGRADUATE TEXTS · 20

Fundamentals of Mathematical Analysis

Paul J. Sally, Jr.

American Mathematical Society
Providence, Rhode Island

EDITORIAL COMMITTEE

Paul J. Sally, Jr. (Chair) Joseph Silverman
Francis Su Susan Tolman

2010 *Mathematics Subject Classification.* Primary 15–01, 22B05, 26–01, 28–01, 42–01, 43–01, 46–01.

For additional information and updates on this book, visit
www.ams.org/bookpages/amstext-20

Library of Congress Cataloging-in-Publication Data

Sally, Paul J., Jr., 1933–
Fundamentals of mathematical analysis / Paul J. Sally, Jr.
pages cm. — (Pure and applied undergraduate texts; volume 20)
Includes bibliographical references and index.
ISBN 978-0-8218-9141-4 (alk. paper)
1. Mathematical analysis. I. Title.

QA300.S23 2013
515—dc23

2012036735

Copying and reprinting. Individual readers of this publication, and nonprofit libraries acting for them, are permitted to make fair use of the material, such as to copy a chapter for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgment of the source is given.

Republication, systematic copying, or multiple reproduction of any material in this publication is permitted only under license from the American Mathematical Society. Requests for such permission should be addressed to the Acquisitions Department, American Mathematical Society, 201 Charles Street, Providence, Rhode Island 02904-2294 USA. Requests can also be made by e-mail to `reprint-permission@ams.org`.

© 2013 by the author.
The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 18 17 16 15 14 13

To my wife and fellow author, Judith D. Sally

Contents

Preface ix

Acknowledgments xiii

Chapter 1. The Construction of Real and Complex Numbers 1
1.1. The Least Upper Bound Property and the Real Numbers 2
1.2. Consequences of the Least Upper Bound Property 4
1.3. Rational Approximation 5
1.4. Intervals 8
1.5. The Construction of the Real Numbers 9
1.6. Convergence in $\mathbb{R}$ 13
1.7. Automorphisms of Fields 17
1.8. Complex Numbers 19
1.9. Convergence in $\mathbb{C}$ 20
1.10. Independent Projects 24

Chapter 2. Metric and Euclidean Spaces 33
2.1. Introduction 34
2.2. Definition and Basic Properties of Metric Spaces 34
2.3. Topology of Metric Spaces 36
2.4. Limits and Continuous Functions 44
2.5. Absolute Continuity and Bounded Variation in $\mathbb{R}$ 51
2.6. Compactness, Completeness, and Connectedness 56
2.7. Independent Projects 64

Chapter 3. Complete Metric Spaces 77
3.1. The Contraction Mapping Theorem and Its Applications to Differential and Integral Equations 78
3.2. The Baire Category Theorem and the Uniform Boundedness Principle 79
3.3. Stone-Weierstrass Theorem 82
3.4. The p-adic Completion of $\mathbb{Q}$ 85
3.5. Independent Projects 93

Chapter 4. Normed Linear Spaces 101
4.1. Definitions and Basic Properties 102
4.2. Bounded Linear Operators 106
4.3. Fundamental Theorems about Linear Operators 109
4.4. Extending Linear Functionals 112
4.5. Generalized Limits and the Dual of $\ell^\infty(F)$ 114
4.6. Adjoint Operators and Isometries of Normed Linear Spaces 116
4.7. Concrete Facts about Isometries of Normed Linear Spaces 119
4.8. Locally Compact Groups 123
4.9. Hilbert Spaces 126
4.10. Convergence and Selfadjoint Operators 134
4.11. Independent Projects 136

Chapter 5. Differentiation 141
5.1. Review of Differentiation in One Variable 142
5.2. Differential Calculus in $\mathbb{R}^n$ 149
5.3. The Derivative as a Matrix of Partial Derivatives 154
5.4. The Mean Value Theorem 158
5.5. Higher-Order Partial Derivatives and Taylor's Theorem 160
5.6. Hypersurfaces and Tangent Hyperplanes in $\mathbb{R}^n$ 164
5.7. Max-Min Problems 166
5.8. Lagrange Multipliers 170
5.9. The Implicit and Inverse Function Theorems 175
5.10. Independent Projects 182

Chapter 6. Integration 191
6.1. Measures 192
6.2. Lebesgue Measure 194
6.3. Measurable Functions 205
6.4. The Integral 208
6.5. L^p Spaces 216
6.6. Fubini's Theorem 220
6.7. Change of Variables in Integration 224
6.8. Independent Projects 227

Chapter 7. Fourier Analysis on Locally Compact Abelian Groups 237
7.1. Fourier Analysis on the Circle 238
7.2. Fourier Analysis on Locally Compact Abelian Groups 246
7.3. The Determination of $\hat{G}$ 249
7.4. The Fourier Transform on $(\mathbb{R}, +)$ 251
7.5. Fourier Inversion on $(\mathbb{R}, +)$ 254
7.6. Fourier Analysis on p-adic Fields 259
7.7. Independent Projects 263

Appendix A. Sets, Functions, and Other Basic Ideas 271
A.1. Sets and Elements 271
A.2. Equality, Inclusion, and Notation 272
A.3. The Algebra of Sets 273
A.4. Cartesian Products, Counting, and Power Sets 277
A.5. Some Sets of Numbers 279
A.6. Equivalence Relations and the Construction of $\mathbb{Q}$ 284
A.7. Functions 290
A.8. Countability and Other Basic Ideas 297
A.9. The Axiom of Choice 305
A.10. Independent Projects 308

Appendix B. Linear Algebra 313
B.1. Fundamentals of Linear Algebra 313
B.2. Linear Transformations 319
B.3. Linear Transformations and Matrices 321
B.4. Determinants 324
B.5. Geometric Linear Algebra 332
B.6. Independent Projects 340

Bibliography 351

Index of Terminology 353

Index of Notation Definitions 361

Preface

What happens when a student completes a course in "AP Calculus", followed by a routine course in multivariable calculus, a computational course in linear algebra, and a formulaic presentation of differential equations? It is time for some real mathematics. There is still a world of interconnected labyrinths to explore, unscalable mountains to be climbed, and lands of mystery to be discovered. The first step in this is a rigorous course in one-variable analysis. This begins with a study of an ordered field in which the least upper bound property holds (not to be confused with a complete ordered field; see the project in Section 2.7.3). Here the student meets Bolzano-Weierstrass, Heine-Borel, and a rigorous treatment of one-variable differentiation and integration with careful attention paid to the pervasive presence of the Mean Value Theorem. Then what? That is exactly the reason behind this book.

Learning serious mathematics is about engaging with problems, from kindergarten to graduate school and beyond. The preliminaries for reading this book are already contained in the author's book *Tools of the Trade* [**27**]. The first two chapters of *Tools* are Appendices A and B of this book. The reader who is familiar with that material can jump right into Chapter 1. The sequence of topics can be gleaned from the table of contents, so I will not dwell on that.

There are three features here that should be discussed explicitly, especially since they are important for the use of this book as a text. The first, and most important, is the collection of exercises. These are spread throughout the chapters and should be regarded as an essential component of the student's learning. Some of these exercises comprise a routine follow-up to the material, while others will challenge the student's understanding more deeply. The second feature is the set of independent projects presented at the end of each chapter. These projects supplement the content studied in their respective chapters. They can be used to expand the student's knowledge and understanding or as an opportunity to conduct a seminar in Inquiry Based Learning (IBL) in which the students present the material to their class. A brief glance will show that the independent projects cover a wide range of interesting topics that hint at advanced areas of mathematics. The

third feature is the real kicker in this business. We list a series of challenge problems that increase in impossibility as the chapters progress.

I have taught the material in this book many times over the past forty-five years. The main audience has been the students in Honors Analysis (MATH 207-208-209) at the University of Chicago. These students are drawn from two sources. The first is the collection of sophomores who have excelled at Honors Calculus in their first year at Chicago. The second is a selection of pyrotechnically endowed freshmen who are capable of attacking mathematics at this level.

Some of the texts I have recommended during this time are T. Apostol, *Mathematical Analysis* [**2**], J. Dieudonné , *Foundations of Modern Analysis* [**3**], A. Kolmogorov and S. Fomin, *Introductory Real Analysis* [**12**], S. Lang, *Undergraduate Analysis* [**13**], L. Loomis and S. Sternberg, *Advanced Calculus* [**18**], W. Rudin, *Principles of Mathematical Analysis* [**24**], and, more recently, C. Pugh, *Real Mathematical Analysis* [**22**]. All of these books have some nice features. The intersection with the material of the present book is highly nontrivial. Nonetheless, I have always liked the idea of challenge problems, independent projects, and the organization of the mathematics presented here. For example, it is about time that mathematicians came to grips with Fourier analysis on p-adic fields, since it is an integral part of current-day research.

At the beginning of each chapter, I have included a quote from a well-known mathematician (or group of mathematicians) that gives a certain perspective on the material in that particular chapter. We leave it to the reader to speculate as to whether this perspective is that of the author. These quotes express a variety of opinions, and I have found them to be informative and sometimes amusing. The quote of A. Zygmund at the beginning of Chapter 7 is particularly relevant to the mathematics in the text.

FURTHER ADVICE TO THE STUDENT

(If you do not care about advice, just get started with the challenge problems in Chapter 1.)

It would be much better for both of us if I were sitting on a desk at the front of the class and talking to you. Nevertheless, a few words of warning are in order. First of all, you should scan the material in Appendices A and B and make sure you feel comfortable with it. Throughout the text, there are many references to these appendices. Secondly, if you find a particular exercise in the text to be quite simple and the next exercise to be very difficult, that's just the way it is. When doing mathematics, you never know when a road that seems smooth is going to have a pothole that is ten feet deep. Thirdly, if you take the challenge problems seriously, you will find that some of these problems can require looking somewhere other than Wikipedia. In that process, you can discover that there is lots of good stuff

in libraries. In any mathematics course of consequence, students should always be willing and ready to find other approaches to the proofs and solutions that are given "in class". In most cases, after Chapter 1, the challenge problems are related to material that has been covered earlier in the text. Good luck and enjoy.

Paul J. Sally, Jr.
Chicago, Illinois
July 31, 2012

Acknowledgments

We acknowledge with gratitude the contributions to this book made by our colleagues, students, and friends over the past forty-five years. It is not possible to mention by name all those with whom we have had conversations about this material, but we express appreciation to them here.

We thank Robert Fefferman, Bill Fulton, Denis Hirschfeldt, Carlos Kenig, David Leep, Raghavan Narasimhan, Madhav Nori, Joseph Silverman, and David Vogan.

We are grateful to Sam Altschul, David Coley, Moon Duchin, Jonny Gleason, Sam Isaacson, Sean Johnson, Ryan Julian, Tom Koberda, Calvin Lin, Chris Malon, Mike Miller, Emily Peters, Sam Raskin, Ryan Reich, Weston Ungemach, Michael Wong, and Alex Zorn for their incisive ideas on the material in this book, as well as their perceptive reading of portions of it.

We are in great debt to Mitya Boyarchenko who contributed significant ideas to the book and to John Boller, Evan Jenkins, and Loren Spice, all of whom also contributed ideas to the book and carefully read versions of the manuscript.

My ultimate debt is owed to those who worked with me to produce this manuscript. The word colleague describes them appropriately. The word amanuensis could be used as a formal title, but they are much more. We argued, discussed, rewrote, reaffirmed, and readjusted parts of the manuscript on many occasions. These friends are Chris Jeris, Nick Ramsey, Kaj Gartz, Nick Ramsey (again), Nick Longo, and finally Nick Ramsey a third time. Working with Nick Ramsey was an especially rewarding experience, since he grew from an undergraduate student in my Honors Analysis class to a graduate student and then to a research mathematician. This book would never have appeared without his cooperation.

Paul J. Sally, Jr.
Chicago, Illinois
July 31, 2012

Chapter 1

The Construction of Real and Complex Numbers

> Thus the System of Real Numbers—the definition of irrationals and the extension of the four species to the new numbers—is established. The method has the advantage of simplicity in detail. It is well for the student, after a first study of the method of Dedekind, to work it through in detail. He will then return to the former method with increased power and greater zest.
>
> The method of regular sequences is a middle-of-the-road method. It is an easy way to reach the mountain top. The traveler buys his ticket and takes the funicular. Many people prefer this mode of travel. But some like a stiff climb over rocks and across streams, and such an ascent has its advantages if the heart is good and the muscles are strong.
>
> – *William Fogg Osgood*
> *Functions of Real Variables*

Challenge Problems: I

PROBLEM 1.1. How many congruent regular tetrahedra with edge length 1 can be packed inside a sphere of radius 1 if each tetrahedron has a vertex at the center of the sphere?

A lattice point in n-dimensional Euclidean space is a point all of whose coordinates are integers. A lattice polygon in $\mathbb{R}^2$ is a polygon all of whose vertices are lattice points. Similarly, a lattice polyhedron in $\mathbb{R}^n$ is a polyhedron all of whose vertices are lattice points.

PROBLEM 1.2. Find all possible areas of lattice squares in $\mathbb{R}^2$. (Hint: The first answer you get is probably not the final answer we are looking for.)

PROBLEM 1.3. Find all possible volumes of lattice cubes in $\mathbb{R}^3$.

PROBLEM 1.4. Find all possible volumes of lattice hypercubes in $\mathbb{R}^n$ for $n > 3$.

PROBLEM 1.5. If $f : \mathbb{R} \to \mathbb{R}$ is a polynomial function such that $f(\mathbb{Q}) \subseteq \mathbb{Q}$ and $f(\mathbb{R} \setminus \mathbb{Q}) \subseteq \mathbb{R} \setminus \mathbb{Q}$, show that $f(x) = ax + b$ for some $a, b \in \mathbb{Q}$.

To begin this chapter, we assume that the reader is familiar with the integers $\mathbb{Z}$ as an ordered integral domain and the rational numbers $\mathbb{Q}$ as an ordered field that is the field of fractions of the integers $\mathbb{Z}$ (see Appendix A).

EXERCISE 1.0.1. Prove that any ordered integral domain contains the integers.

EXERCISE 1.0.2. Prove that any field that contains the integers contains the rationals as a subfield.

In this chapter, we do several things. First, we introduce the real numbers by adding the least upper bound property to the axioms for an ordered field. Second, despite Osgood, we construct the real numbers from the rational numbers by the method of Cauchy sequences. Third, we construct the complex numbers from the real numbers and prove a few useful theorems about them. Intermingled in all of this is a discussion of the fields of algebraic numbers and real algebraic numbers. As a project at the end of the chapter, we lead the reader through a discussion of the construction of the real numbers via Dedekind cuts. In other projects, we study the convergence properties of infinite series and decimal expansions of real numbers.

1.1. The Least Upper Bound Property and the Real Numbers

DEFINITION 1.1.1. Let F be an ordered field. Let A be a nonempty subset of F. We say that A is *bounded above* if there is an element $M \in F$ with the property that if $x \in A$, then $x \leq M$. We call M an *upper bound* for A. Similarly, we say that A is *bounded below* if there is an element $m \in F$ such that if $x \in A$, then $m \leq x$. We call m a *lower bound* for A. We say that A is *bounded* if A is bounded above and A is bounded below.

EXAMPLES 1.1.2.

(i) Consider the subset A of $\mathbb{Q}$:

$$A = \left\{ 1 + \frac{(-1)^n}{n} \;\middle|\; n \in \mathbb{N} \right\}.$$

Then A is bounded above by $3/2$ and bounded below by 0.

(ii) Let $A = \{x \in \mathbb{Q} \mid 0 < x^3 < 27\}$. Then A is bounded below by 0 and bounded above by 3.

EXERCISE 1.1.3. Let a be a positive rational number and let $A = \{x \in \mathbb{Q} \mid x^2 < a\}$. Show that A is bounded in $\mathbb{Q}$.

DEFINITION 1.1.4. Let F be an ordered field, and let A be a nonempty subset of F which is bounded above. We say that $L \in F$ is a *least upper bound* for A if the following two conditions hold:

(a) L is an upper bound for A;
(b) if M is any upper bound for A, then $L \leq M$.

EXERCISE 1.1.5. Show the least upper bound of a set is unique.

We now give a formal definition of the real numbers which provides a working basis for proving theorems. Later in this chapter, starting with the rational numbers as an ordered field we will give a precise construction of the real numbers as an ordered field in which the least upper bound property holds.

DEFINITION 1.1.6. The *real numbers* are an ordered field in which every nonempty subset that is bounded above has a least upper bound and are denoted by the symbol $\mathbb{R}$.

We say that the real numbers are an ordered field with the least upper bound property. In many texts, the real numbers are defined as a *complete ordered field*. This is actually a misuse of the word "complete" which is defined in terms of the convergence of Cauchy sequences. This will be discussed later in this chapter.

EXERCISE 1.1.7. Find the least upper bound in $\mathbb{R}$ of the set A in Exercise 1.1.3.

DEFINITION 1.1.8. Suppose that F and F' are ordered integral domains. We say that F and F' are *order isomorphic* if there is a bijection $\phi : F \longrightarrow F'$ such that

(a) $\phi(x + y) = \phi(x) + \phi(y)$ for all $x, y \in F$;
(b) $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in F$;
(c) if $x, y \in F$ and $x < y$, then $\phi(x) < \phi(y)$ in F'.

EXERCISE 1.1.9. Show that any two ordered fields with the least upper bound property are order isomorphic.

This exercise proves that if the real numbers exist, they are unique up to order isomorphism.

DEFINITION 1.1.10. An ordered field F has the *greatest lower bound property* if every nonempty subset A of F that is bounded below has a greatest lower bound. That is, there exists an element ℓ of F such that:

(a) ℓ is a lower bound for A;
(b) if m is any lower bound for A, then $m \leq \ell$.

EXERCISE 1.1.11. Prove that an ordered field has the least upper bound property iff it has the greatest lower bound property.

If L is the least upper bound of a set A, we write $L = \text{lub } A$ or $L = \sup A$ (sup stands for *supremum*). If ℓ is the greatest lower bound of a set A, we write $\ell = \text{glb } A$ or $\ell = \inf A$ (inf stands for *infimum*).

EXERCISE 1.1.12. Let n be a positive integer that is not a perfect square. Let $A = \{x \in \mathbb{Q} \mid x^2 < n\}$. Show that A is bounded in $\mathbb{Q}$ but has neither a greatest lower bound nor a least upper bound in $\mathbb{Q}$. Conclude that $\sqrt{n}$ exists in $\mathbb{R}$, that is, there exists a real number a such that $a^2 = n$.

We have observed that the rational numbers are contained in $\mathbb{R}$. A real number is *irrational* if it is not in $\mathbb{Q}$.

FACT 1.1.13. We can conclude from Exercise 1.1.12 that if n is a positive integer that is not a perfect square, then $\sqrt{n}$ exists in $\mathbb{R}$ and is irrational.

EXERCISE 1.1.14. Suppose that A and B are bounded sets in $\mathbb{R}$. Prove or disprove the following:

(i) $\text{lub}(A \cup B) = \max\{\text{lub } A, \text{lub } B\}$.
(ii) If $A + B = \{a + b \mid a \in A, b \in B\}$, then $\text{lub}(A + B) = \text{lub } A + \text{lub } B$.
(iii) If the elements of A and B are positive and $A \cdot B = \{ab \mid a \in A, b \in B\}$, then $\text{lub}(A \cdot B) = (\text{lub } A)(\text{lub } B)$.
(iv) Formulate the analogous problems for the greatest lower bound.

1.2. Consequences of the Least Upper Bound Property

We now present some facts which follow from the least upper bound property and the properties of the integers. The first is the *Archimedean Property* of the real numbers.

THEOREM 1.2.1 **(Archimedean property of $\mathbb{R}$)**. *If a and b are positive real numbers, then there exists a natural number n such that $na > b$.*

Proof. If $a > b$, take $n = 1$. If $a = b$, take $n = 2$. If $a < b$, consider the set $S = \{na \mid n \in \mathbb{N}\}$. The set $S \neq \varnothing$ since $a \in S$. Suppose S is bounded above by b. Let $L = \text{lub } S$. Then, since $a > 0$, there exists an element $n_0 a \in S$ such that $L - a < n_0 a$. But then $L < (n_0 + 1)a$, which is a contradiction.

COROLLARY 1.2.2. *If ε is a positive real number, then there exists a natural number n such that $1/n < \varepsilon$.*

DEFINITION 1.2.3. Let F be an ordered field. From Exercise 1.0.1, we know that $\mathbb{Z} \subseteq F$ and by Exercise 1.0.2 we know $\mathbb{Q} \subseteq F$. We say that F is an *Archimedean ordered field* if for every $x \in F$ there exists $N \in \mathbb{Z}$ such that $x < N$.

The fields $\mathbb{Q}$ and $\mathbb{R}$ are Archimedean ordered fields.

EXERCISE 1.2.4. Let F be an Archimedean ordered field. Show that F is order isomorphic to a subfield of $\mathbb{R}$.

Next, we show that every real number lies between two successive integers.

THEOREM 1.2.5. *If a is a real number, then there exists an integer N such that $N - 1 \leq a < N$.*

Proof. Let $S = \{n \in \mathbb{Z} \mid n > a\}$. Then by the Archimedean property, $S \neq \varnothing$. The set S is bounded below by a, so by the well-ordering principle, S has a least element N. Then $N - 1 \notin S$, so $N - 1 \leq a < N$.

We now show that there is a rational number between any two real numbers.

THEOREM 1.2.6. *If a and b are real numbers with $a < b$, then there exists a rational number $r = p/q$ such that $a < r < b$.*

Proof. From the Archimedean property of $\mathbb{R}$ (Corollary 1.2.2) there exists $q \in \mathbb{N}$ such that $1/q < b - a$. Now consider the real number qa. By Theorem 1.2.5, there exists an integer p such that $p - 1 \leq qa < p$. It follows that $\frac{p-1}{q} \leq a < \frac{p}{q}$. This implies that $\frac{p}{q} - \frac{1}{q} \leq a$, that is, $a < \frac{p}{q} \leq a + \frac{1}{q} < b$.

DEFINITION 1.2.7. A subset A of $\mathbb{R}$ is said to be *dense* in $\mathbb{R}$ if for any pair of real numbers a and b with $a < b$, there is an $r \in A$ such that $a < r < b$.

COROLLARY 1.2.8. *The rational numbers are dense in the real numbers.*

How do the irrational numbers behave?

EXERCISE 1.2.9.

(i) Show that any irrational number multiplied by any nonzero rational number is irrational.

(ii) Show that the product of two irrational numbers may be rational or irrational.

Next we show that there is an irrational number between any two real numbers.

COROLLARY 1.2.10. *The irrational numbers are dense in $\mathbb{R}$.*

Proof. Take $a, b \in \mathbb{R}$ such that $a < b$. We know that $\sqrt{2}$ is irrational and greater than 0. But then $\frac{a}{\sqrt{2}} < \frac{b}{\sqrt{2}}$. By Corollary 1.2.8, there exists a rational number p/q with $p \neq 0$ such that $\frac{a}{\sqrt{2}} < \frac{p}{q} < \frac{b}{\sqrt{2}}$. Thus $a < \sqrt{2}p/q < b$, and $\sqrt{2}p/q$ is irrational.

The real numbers are the union of two disjoint sets, the rational numbers and the irrational numbers, and each of these sets is dense in $\mathbb{R}$. Density implies nothing about cardinality since the rationals are countable and the irrationals are not, as shown in Appendix A.

1.3. Rational Approximation

We have just shown that both the rational numbers and the irrational numbers are dense in the real numbers. But, really, how dense are they? It is reasonable to think that proximity for rational numbers can be measured in terms of the size of the denominator. To illustrate this, we ask the question, "How close do two rational numbers have to be in order to be the

same rational number?" This is not a trick question—it is designed to illustrate the principle mentioned above. In particular, if $a/b, c/d \in \mathbb{Q}$ and $|a/b - c/d| < 1/bd$, then $a/b = c/d$.

This idea can be encapsulated in the following theorem. Throughout this section, we shall assume that the denominator of a rational number is a positive integer and that the numerator and denominator are relatively prime.

THEOREM 1.3.1. *If a/b is a fixed rational number and p/q is a rational number such that $0 < |p/q - a/b| < 1/mb$ for some positive integer m, then $q > m$.*

Proof. Easy.

We now present several facts on rational approximation. For α in various subsets of the real numbers, we prove results which give an idea of the degree of accuracy with which α may be approximated. The results take the following form: Given a real number α,

(1) positive real numbers $c(\alpha)$ and t exist so that there are infinitely many rational numbers p/q with $|\alpha - p/q| < c(\alpha)/q^t$;
(2) positive real numbers $c(\alpha)$ and t exist so that there are only finitely many rational numbers p/q with $|\alpha - p/q| < c(\alpha)/q^t$;
(3) for $\delta > 0$, there exist real numbers $c(\alpha, \delta)$ and t so that $|\alpha - p/q| \geq c(\alpha, \delta)/q^{t+\delta}$ for all rational numbers p/q.

To begin, we present an exercise which follows easily from elementary number theory.

EXERCISE 1.3.2. Let a and b be relatively prime integers. Show that the equation $ax + by = 1$ has infinitely many solutions (x, y) with x and y relatively prime.

THEOREM 1.3.3. *Let $\alpha = a/b$ with a and b relatively prime and $b \neq 1$. Then there exist infinitely many $p/q \in \mathbb{Q}$ such that $|a/b - p/q| < 1/q$.*

Proof. Let $(x, y) = (q, -p)$ be a solution to the equation $ax + by = 1$. Then $q \neq 0$ since $b \neq 1$. We may assume $q > 0$. We then have $|a/b - p/q| = 1/bq < 1/q$.

REMARK 1.3.4. If $b = 1$, then the same result holds with $<$ replaced by $\leq$.

The next theorem characterizes rational numbers in terms of rational approximation. We first need the following exercise.

EXERCISE 1.3.5. Let α be a real number, and let η and t be positive real numbers. Show that there exists only a finite number of rational numbers p/q with $q < \eta$ which satisfy $|\alpha - p/q| < 1/q^t$.

THEOREM 1.3.6. *Let $\alpha = a/b \in \mathbb{Q}$. Then there are only finitely many p/q so that $|a/b - p/q| \leq 1/q^2$.*

Proof. Suppose there are infinitely many p/q satisfying the inequality. Then by the exercise above, q gets arbitrarily large. Thus there exists a p/q with $q > b$ such that $|a/b - p/q| < 1/q^2$. This implies that $|aq - bp| < b/q < 1$, which is a contradiction.

We next consider rational approximation of irrational numbers. The question is, "If α is irrational, are there any rational numbers p/q satisfying the inequality $|\alpha - p/q| < 1/q^2$?" The affirmative answer follows from a theorem of Dirichlet on rational approximation of any real number.

THEOREM 1.3.7 **(Dirichlet)**. *Let α be a real number and n a positive integer. Then there is a rational number p/q with $0 < q \leq n$ satisfying the inequality*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{(n+1)q}.$$

Proof. If $n = 1$, then $p/q = [\alpha]$ or $p/q = [\alpha + 1]$ satisfies $|\alpha - p/q| \leq 1/2$. Suppose that $n \geq 2$. Consider the $n + 2$ numbers

$$0, \alpha - [\alpha], 2\alpha - [2\alpha], \ldots, n\alpha - [n\alpha], 1$$

in the interval $[0, 1]$. Assume that the numbers in our list are distinct, which is the case if α is irrational. By the pigeonhole principle, two of the numbers differ in absolute value by at most $1/(n+1)$. If one of the numbers is 0 and the other is $i\alpha - [i\alpha]$, then $i \leq n$, $|i\alpha - [i\alpha]| \leq 1/(n+1)$, and

$$\left|\alpha - \frac{[i\alpha]}{i}\right| \leq \frac{1}{(n+1)i}.$$

After $[i\alpha]/i$ is reduced to lowest terms p/q, the rational number p/q satisfies the required inequality. Similarly, if the two numbers are $j\alpha - [j\alpha]$ and 1, then $j \leq n$, and reducing $([j\alpha] + 1)/j$ to lowest terms p/q, we have p/q satisfies the required inequality. Finally, if the two numbers are $i\alpha - [i\alpha]$ and $j\alpha - [j\alpha]$, where $i < j$, then

$$|j\alpha - [j\alpha] - (i\alpha - [i\alpha])| = |(j-i)\alpha + ([j\alpha] - [i\alpha])| \leq \frac{1}{n+1}.$$

Then, $j - i < n$, and

$$\left|\alpha - \frac{[j\alpha] - [i\alpha]}{j - i}\right| \leq \frac{1}{(n+1)(j-i)}.$$

Thus, after $([j\alpha] - [i\alpha])/(j - i)$ is reduced to lowest terms p/q, the rational number p/q satisfies the inequality. In the event that the $n+2$ numbers are not distinct, then α itself is a rational number with denominator at most n. For this case, either there exists $1 \leq i \leq n$ so that

$$\alpha = \frac{[i\alpha]}{i}$$

or there exist $1 \leq i < j \leq n$ so that

$$\alpha = \frac{[j\alpha] - [i\alpha]}{j - i}.$$

Thus, if the numbers are not distinct, the required inequality is trivially satisfied by α itself.

COROLLARY 1.3.8. *Given any real number α, there is a rational number p/q such that $|\alpha - p/q| < 1/q^2$.*

Proof. This follows immediately from the theorem.

Now comes the good news (or bad news depending on how you look at it).

THEOREM 1.3.9. *If α is irrational, then there are infinitely many rational numbers p/q such that $|\alpha - p/q| < 1/q^2$.*

Proof. Suppose there are only a finite number of rational numbers p_1/q_1, $p_2/q_2, \ldots, p_k/q_k$ satisfying the inequality. Then, there is a positive integer n such that $|\alpha - p_i/q_i| > 1/(n+1)q_i$ for $i = 1, 2, \ldots, k$. This contradicts Theorem 1.3.7, which asserts the existence of a rational number p/q satisfying $q \leq n$ and $|\alpha - p/q| < 1/(n+1)q < 1/q^2$.

So, there you have it, a real number α is rational if and only if there exists only a finite number of rational numbers p/q such that $|\alpha - p/q| \leq 1/q^2$. Moreover, a real number α is irrational if and only if there exists an infinite number of rational numbers p/q such that $|\alpha - p/q| \leq 1/q^2$.

1.4. Intervals

At this stage we single out certain subsets of $\mathbb{R}$ which are called intervals.

DEFINITION 1.4.1. A subset of $\mathbb{R}$ is an *interval* if it falls into one of the following categories.

(a) For $a, b \in \mathbb{R}$ with $a < b$, the *open interval* (a, b) is defined by $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$.
(b) For $a, b \in \mathbb{R}$ with $a \leq b$, the *closed interval* $[a, b]$ is defined by $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$.
(c) For $a, b \in \mathbb{R}$ with $a < b$, the *half-open interval* $[a, b)$ is defined by $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$.
(d) For $a, b \in \mathbb{R}$ with $a < b$, the *half-open interval* $(a, b]$ is defined by $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$.
(e) For $a \in \mathbb{R}$, the *infinite open interval* (a, ∞) is defined by $(a, \infty) = \{x \in \mathbb{R} \mid a < x\}$.
(f) For $b \in \mathbb{R}$, the *infinite open interval* $(-\infty, b)$ is defined by $(-\infty, b) = \{x \in \mathbb{R} \mid x < b\}$.
(g) For $a \in \mathbb{R}$, the *infinite closed interval* $[a, \infty)$ is defined by $[a, \infty) = \{x \in \mathbb{R} \mid a \leq x\}$.

(h) For $b \in \mathbb{R}$, the *infinite closed interval* $(-\infty, b]$ is defined by $(-\infty, b] = \{x \in \mathbb{R} \mid x \leq b\}$.
(i) $\mathbb{R} = (-\infty, \infty)$.

DEFINITION 1.4.2. If $x \in \mathbb{R}$, a *neighborhood* of x is an open interval containing x. In many instances, it is helpful to use symmetric neighborhoods. That is, if $x \in \mathbb{R}$, a *symmetric neighborhood* of x is an interval of the form $(x - \varepsilon, x + \varepsilon)$, where $\varepsilon > 0$.

These intervals, and their counterparts in other spaces, are used extensively throughout analysis.

EXERCISE 1.4.3. Suppose that I is a subset of $\mathbb{R}$. Show that I is an interval if and only if for all $a, b \in I$, with $a \leq b$, the closed interval $[a, b] \subseteq I$.

The notion of interval is valid in any ordered field, and we will occasionally find this useful. We end this section with a theorem about intervals in $\mathbb{R}$, which is called the Nested Intervals Theorem.

THEOREM 1.4.4 **(Nested Intervals Theorem)**. *Let $([a_n, b_n])_{n\in\mathbb{N}}$ be a nested sequence of closed bounded intervals in $\mathbb{R}$. That is, for any n we have $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$, or equivalently, $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$ for all n. Then $\bigcap_{n\in\mathbb{N}}[a_n, b_n] \neq \varnothing$.*

Proof. Let $A = \{a_n \mid n \in \mathbb{N}\}$. Then A is bounded above by b_1. If $a = \text{lub}A$, then $a \in \bigcap_{n\in\mathbb{N}}[a_n, b_n]$.

The nested intervals property is actually not exclusive to the real numbers. In fact, it is really a theorem about a sequence of nested compact sets in a metric space. This result will be proved in the next chapter. There is often some confusion about the relationship between the Nested Interval Theorem in $\mathbb{R}$ and the least upper bound property. Although our proof in $\mathbb{R}$ involves the least upper bound property, it can be done in alternate ways which involve sequential compactness.

1.5. The Construction of the Real Numbers

We are now ready to proceed with the construction of the real numbers from the rational numbers using the fact that the rational numbers are the field of fractions of $\mathbb{Z}$. We have already defined $\mathbb{R}$ as an ordered field in which the least upper bound property holds. We now proceed to build such a field starting from $\mathbb{Q}$.

Recall that the *absolute value* on $\mathbb{Q}$ is defined as

$$|a| = \begin{cases} a & \text{if } a \geq 0, \\ -a & \text{if } a < 0. \end{cases}$$

Also recall that the absolute value on $\mathbb{Q}$ satisfies the following three properties.

(1) For any $a \in \mathbb{Q}$, $|a| \geq 0$, and $|a| = 0$ if and only if $a = 0$.
(2) For any $a, b \in \mathbb{Q}$, $|ab| = |a||b|$.
(3) For any $a, b \in \mathbb{Q}$, $|a + b| \leq |a| + |b|$ (triangle inequality).

EXERCISE 1.5.1. Show that, for any $a, b \in \mathbb{Q}$, we have $||a| - |b|| \leq |a - b|$.

DEFINITION 1.5.2. A sequence $(a_k)_{k\in\mathbb{N}}$ of rational numbers is a *Cauchy sequence* in $\mathbb{Q}$ if, given any rational number $r > 0$, there exists an integer N such that if $n, m \geq N$, then $|a_n - a_m| < r$.

DEFINITION 1.5.3. A sequence $(a_k)_{k\in\mathbb{N}}$ *converges in* $\mathbb{Q}$ to $a \in \mathbb{Q}$ if, given any rational number $r > 0$, there exists an integer N such that if $n \geq N$, then $|a_n - a| < r$. Sometimes, we just say that the sequence $(a_k)_{k\in\mathbb{N}}$ converges in $\mathbb{Q}$ without mentioning the *limit* a.

EXERCISE 1.5.4. If a sequence $(a_k)_{k\in\mathbb{N}}$ converges in $\mathbb{Q}$, show that $(a_k)_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{Q}$. In addition, show also that the limit a of a convergent sequence is unique.

DEFINITION 1.5.5. Let $(a_k)_{k\in\mathbb{N}}$ be a sequence of rational numbers. We say that $(a_k)_{k\in\mathbb{N}}$ is a *bounded sequence* if the set $\{a_k \mid k \in \mathbb{N}\}$ is a bounded set in $\mathbb{Q}$.

LEMMA 1.5.6. *Let $(a_k)_{k\in\mathbb{N}}$ be a Cauchy sequence of rational numbers. Then $(a_k)_{k\in\mathbb{N}}$ is a bounded sequence.*

Proof. Let $(a_k)_{k\in\mathbb{N}}$ be a Cauchy sequence of rational numbers. Pick $N \in \mathbb{N}$ such that $|a_n - a_m| < 1$ for $n, m \geq N$. Then $|a_n - a_N| < 1$ for all $n \geq N$, so that $|a_n| < 1 + |a_N|$ for all $n \geq N$. Let M be the max of $|a_1|, |a_2|, \ldots, |a_{N-1}|, 1 + |a_N|$. Then $(|a_k|)_{k\in\mathbb{N}}$ is bounded by M.

Let $\mathcal{C}$ denote the set of all Cauchy sequences of rational numbers. We define addition and multiplication of Cauchy sequences termwise; that is, $(a_k)_{k\in\mathbb{N}} + (b_k)_{k\in\mathbb{N}} = (a_k + b_k)_{k\in\mathbb{N}}$ and $(a_k)_{k\in\mathbb{N}}(b_k)_{k\in\mathbb{N}} = (a_k b_k)_{k\in\mathbb{N}}$.

EXERCISE 1.5.7. Show that the sum of two Cauchy sequences in $\mathbb{Q}$ is a Cauchy sequence in $\mathbb{Q}$.

THEOREM 1.5.8. *The product of two Cauchy sequences in $\mathbb{Q}$ is a Cauchy sequence in $\mathbb{Q}$.*

Proof. Let $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ be Cauchy sequences in $\mathbb{Q}$. Then

$$\begin{aligned} |a_n b_n - a_m b_m| &= |a_n b_n - a_n b_m + a_n b_m - a_m b_m| \\ &\leq |a_n||b_n - b_m| + |b_m||a_n - a_m| \\ &\leq A|b_n - b_m| + B|a_n - a_m|, \end{aligned}$$

where A and B are upper bounds for the sequences $(|a_k|)_{k\in\mathbb{N}}$ and $(|b_k|)_{k\in\mathbb{N}}$. Since $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ are Cauchy sequences, the theorem now follows.

EXERCISE 1.5.9. Show that, with addition and multiplication defined as above, $\mathcal{C}$ is a commutative ring with 1 (see Appendix A).

Now let $\mathcal{I}$ be the subset of $\mathcal{C}$ consisting of sequences $(a_k)_{k\in\mathbb{N}}$ with the property that, given any rational $r > 0$, there exists an integer N such that if $n \geq N$, then $|a_n| < r$. The set $\mathcal{I}$ consists of Cauchy sequences which converge to 0.

Suppose $(a_k)_{k\in\mathbb{N}} \notin \mathcal{I}$. Then there exists an $r > 0$ such that $|a_k| \geq r$ infinitely often. Pick $N \in \mathbb{N}$ such that $|a_n - a_m| < r/2$ for $n, m \geq N$. This implies that

$$|a_n| > |a_m| - \frac{1}{2}r \quad \text{for } n, m \geq N.$$

Fix an $m \geq N$ for which $|a_m| \geq r$. Then for all $n \geq N$, we have

$$|a_n| > \frac{1}{2}r.$$

Thus, Cauchy sequences which do not converge to 0 are eventually bounded below (in absolute value) by some positive constant.

EXERCISE 1.5.10. Show that if a Cauchy sequence does not converge to 0, all the terms of the sequence eventually have the same sign.

DEFINITION 1.5.11. Let $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ be Cauchy sequences in $\mathbb{Q}$. We say that $(a_k)_{k\in\mathbb{N}}$ is *equivalent to* $(b_k)_{k\in\mathbb{N}}$, denoted by $(a_k)_{k\in\mathbb{N}} \sim (b_k)_{k\in\mathbb{N}}$, if $(c_k)_{k\in\mathbb{N}} = (a_k - b_k)_{k\in\mathbb{N}}$ is in $\mathcal{I}$.

EXERCISE 1.5.12. Show that $\sim$ defines an equivalence relation on $\mathcal{C}$.

EXERCISE 1.5.13. Given $a \in \mathbb{Q}$, show that the collection of Cauchy sequences in $\mathcal{C}$ converging to a is an equivalence class. In particular, $\mathcal{I}$ is an equivalence class.

Denote by $\mathbf{R}$ the set of equivalence classes in $\mathcal{C}$. We claim that, with appropriate definitions of addition and multiplication (already indicated above) and order (to be defined below), $\mathbf{R}$ is an ordered field satisfying the least upper bound property.

If $(a_k)_{k\in\mathbb{N}}$ is a Cauchy sequence, denote its equivalence class by $[a_k]$. As one might expect, the sum and product of equivalence classes are defined as follows: $[a_k] + [b_k] = [a_k + b_k]$ and $[a_k][b_k] = [a_k b_k]$.

EXERCISE 1.5.14. Show that addition and multiplication are well-defined on $\mathbf{R}$.

EXERCISE 1.5.15. Show that $\mathbf{R}$ is a commutative ring with 1, with $\mathcal{I}$ as the additive identity and $[a_k]$ such that $a_k = 1$ for all k as the multiplicative identity. This follows easily from Exercise 1.5.9.

THEOREM 1.5.16. **R** *is a field.*

Proof. We need only show that multiplicative inverses exist for nonzero elements. So assume that $[a_k] \neq \mathcal{I}$. Then, as we saw above, a_k is eventually bounded below in absolute value. Hence, we can pick $M \in \mathbb{N}$ and $c > 0$ such that $|a_k| > c$ for all $k \geq M$. Define a sequence $(b_k)_{k\in\mathbb{N}}$ as follows: $b_k = 1$ for $k \leq M$, and $b_k = 1/a_k$ for $k > M$. Observe that for n, m large enough

$$\left|\frac{1}{a_n} - \frac{1}{a_m}\right| = \frac{|a_n - a_m|}{|a_n a_m|} \leq \frac{|a_n - a_m|}{c^2}.$$

So $(b_k)_{k\in\mathbb{N}}$ is a Cauchy sequence and $[b_k]$ is the multiplicative inverse of $[a_k]$.

The next step is to define order on **R**. Let $[a_k]$ and $[b_k]$ represent distinct elements of **R**. Then $[c_k] = [a_k - b_k]$ is not equal to $\mathcal{I}$. Hence there exists $N \in \mathbb{N}$ such that all the terms of c_k have the same sign for $k > N$. Thus, either $a_k < b_k$ for all $k \geq N$ or $b_k < a_k$ for $k \geq N$. We use this fact to define an order on **R**.

DEFINITION 1.5.17. Let $a = [a_k], b = [b_k]$ be distinct elements of **R**. We define $a < b$ if $a_k < b_k$ eventually and $b < a$ if $b_k < a_k$ eventually.

EXERCISE 1.5.18. Show that the order relation on **R** defined above is well-defined and makes **R** an ordered field.

To finish this, we must show that **R** is an Archimedean ordered field that satisfies the least upper bound property. We will then have reached the mountain top so we can dismount the funicular and ski happily down the slope.

Define a map $i : \mathbb{Q} \longrightarrow \mathbf{R}$ by sending $r \in \mathbb{Q}$ to the equivalence class of $(r, r, \dots)$. It is evident that this map is injective and order-preserving, so we may consider $\mathbb{Q} \subseteq \mathbf{R}$ as ordered fields.

THEOREM 1.5.19. *The field* **R** *is an Archimedean ordered field.*

Proof. Suppose $a \in \mathbf{R}$ and $a > 0$. Let $(a_k)_{k\in\mathbb{N}}$ represent a. As noted above, the Cauchy sequence $(a_k)_{k\in\mathbb{N}}$ is bounded above by some integer N; that is, $a_k < N$ for all sufficiently large k. It follows that a is less than the integer $(N, N, \dots)$ in **R** (under the inclusion $\mathbb{Q} \subseteq \mathbf{R}$).

THEOREM 1.5.20. *The least upper bound property holds in* **R**.

Proof. Let A be a nonempty subset of **R** that is bounded above by, say, m. Then, by the Archimedean property, we can find $M \in \mathbb{Z}$ with $m \leq M$. Let a be in A and let n be an integer with $n < a$. For $p \in \mathbb{N}$ set $S_p = \{k2^{-p} \mid k \in \mathbb{Z} \text{ and } n \leq k2^{-p} \leq M\} \cup \{m\}$. Note that $S_p \neq \varnothing$ and is finite. Now let $a_p = \min\{x \mid x \in S_p \text{ and } x \text{ is an upper bound for } A\}$.

Note that if $p < q$, then

$$a_p - 2^{-p} < a_q \leq a_p,$$

since, for example, $a_p - 2^{-p}$ is not an upper bound for A, while a_q is an upper bound. But this implies that

$$|a_p - a_q| \leq 2^{-p} \quad \text{for all } p < q,$$

from which it follows that $(a_k)_{k\in\mathbb{N}}$ is a Cauchy sequence. Let $L = [a_k]$.

We claim that L is a least upper bound for A. Suppose $x \in A$ and $x > L$. Choose p such that $2^{-p} < (x - L)$ (using the Archimedean property). Since $a_p - 2^{-p} < a_q$ for $p < q$ and $(a_p)_{p\in\mathbb{N}}$ is a decreasing Cauchy sequence, it follows that $a_p - 2^{-p} \leq L \leq a_p$. In particular, if we add $2^{-p} < x - L$ and $a_p - 2^{-p} \leq L$, we obtain $a_p < x$, which is a contradiction. Therefore L is an upper bound for A.

Suppose that H is an upper bound for A and that $H < L$. Choose p such that $2^{-p} < L - H$. Take $x \in A$ such that $a_p - 2^{-p} < x$. Then $a_p - 2^{-p} < H$. Adding, we get $a_p < L$. But, as noted above, $L \leq a_p$ for all $p \in \mathbb{N}$, so this is a contradiction.

EXERCISE 1.5.21. Prove that $\mathbf{R}$ is order-isomorphic to $\mathbb{R}$. (Hint: You have already done this.)

1.6. Convergence in $\mathbb{R}$

We define the absolute value on $\mathbb{R}$ in exactly the same manner as on $\mathbb{Q}$.

DEFINITION 1.6.1. Suppose $x \in \mathbb{R}$. The *absolute value* of x is defined by

$$|x| = \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

The following are the essential properties of the absolute value.

THEOREM 1.6.2 **(Properties of the absolute value on $\mathbb{R}$).**

(1) *For any $x \in \mathbb{R}$, $|x| \geq 0$, and $|x| = 0$ iff $x = 0$.*
(2) *For any $x, y \in \mathbb{R}$, $|xy| = |x||y|$.*
(3) *For any $x, y \in \mathbb{R}$, $|x + y| \leq |x| + |y|$ (triangle inequality).*

EXERCISE 1.6.3. Prove these properties of the absolute value.

With absolute value defined, we can talk about Cauchy and convergent sequences in $\mathbb{R}$.

DEFINITION 1.6.4. A sequence $(a_k)_{k\in\mathbb{N}}$ of real numbers is *convergent* if there exists an element $a \in \mathbb{R}$ such that the sequence satisfies the following property: given any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $k \geq N$ implies that $|a_k - a| < \varepsilon$. We say that $(a_k)_{k\in\mathbb{N}}$ *converges to* a, and a is called the *limit* of the sequence $(a_k)_{k\in\mathbb{N}}$. Symbolically, we write

$$\lim_{k\to\infty} a_k = a.$$

We will often say that a sequence of real numbers is convergent without specific reference to the limit a. Note that N depends on ε.

EXERCISE 1.6.5. Show that the limit a of a convergent sequence is unique.

DEFINITION 1.6.6. A sequence $(a_k)_{k\in\mathbb{N}}$ of real numbers is *monotonic increasing* if $a_k \leq a_{k+1}$ for all $k \in \mathbb{N}$. A sequence $(a_k)_{k\in\mathbb{N}}$ of real numbers is *strictly monotonic increasing* if $a_k < a_{k+1}$ for all $k \in \mathbb{N}$. Monotonic decreasing and strictly monotonic decreasing sequences are defined similarly.

EXERCISE 1.6.7. Define the notion of a bounded sequence in $\mathbb{R}$.

The following lemma is one of the more useful lemmas in discussing convergence in $\mathbb{R}$ (and $\mathbb{R}^n$).

LEMMA 1.6.8. *Let $(a_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}$. Then $(a_k)_{k\in\mathbb{N}}$ has a monotonic subsequence.*

Proof. Suppose $(a_k)_{k\in\mathbb{N}}$ does not have a monotonic increasing subsequence. Then, there exists $n_1 \in \mathbb{N}$ such that $a_{n_1} > a_k$ for all $k > n_1$. Again, since $(a_k)_{k>n_1}$ does not have a monotonic increasing subsequence, there exists $n_2 > n_1$ such that $a_{n_2} > a_k$ for all $k > n_2$. Moreover $a_{n_1} > a_{n_2}$. Continuing in this way, we obtain a strictly monotonic decreasing subsequence.

LEMMA 1.6.9. *Every bounded monotonic sequence converges in $\mathbb{R}$.*

Proof. Suppose $(a_k)_{k\in\mathbb{N}}$ is monotonic increasing and bounded. Let a be the least upper bound of the set $\{a_1, a_2, \ldots\}$. For all $\varepsilon > 0$, there exists an N such that $a - \varepsilon < a_N \leq a$. Since $(a_k)_{k\in\mathbb{N}}$ is increasing, if $k > N$, we have $a \geq a_k \geq a_N > a - \varepsilon$. So $\lim_{k\to\infty} a_k = a$.

The next lemma is basic for analysis on $\mathbb{R}$.

LEMMA 1.6.10. *Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

EXERCISE 1.6.11. Prove Lemma 1.6.10. This should not take long.

The next definition should be compared with Definition 1.5.2.

DEFINITION 1.6.12. A sequence $(a_k)_{k\in\mathbb{N}}$ in $\mathbb{R}$ is a *Cauchy sequence* if, given any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $n, m \geq N$ implies $|a_m - a_n| < \varepsilon$.

EXERCISE 1.6.13.

(i) Prove that every Cauchy sequence in $\mathbb{R}$ is bounded.
(ii) If $(a_k)_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, show that for any $\varepsilon > 0$ there exists a subsequence $(a_j)_{j\in\mathbb{N}}$ so that $|a_j - a_{j+1}| < \varepsilon/2^{j+1}$.

THEOREM 1.6.14 **(Cauchy criterion)**. *A sequence $(a_k)_{k\in\mathbb{N}}$ of real numbers is convergent if and only if it is a Cauchy sequence.*

Proof. We already did half of this in $\mathbb{Q}$, but we will do it again. First, we prove that if $(a_k)_{k\in\mathbb{N}}$ is convergent, then it is Cauchy. Suppose $\lim_{k\to\infty} a_k =$

a. Then, since the sequence converges, given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $|a_n - a| < \frac{\varepsilon}{2}$ for all $n \geq N$. Thus, if $n, m \geq N$, we have

$$|a_n - a_m| \leq |a_n - a| + |a_m - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

and so $(a_k)_{k\in\mathbb{N}}$ is a Cauchy sequence.

Suppose now that $(a_k)_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. Then, by Exercise 1.6.13, $(a_k)_{k\in\mathbb{N}}$ is a bounded sequence and hence by Lemma 1.6.10 has a convergent subsequence. Call the limit of this subsequence a. Then, since $(a_k)_{k\in\mathbb{N}}$ is Cauchy, it is clear that $\lim_{k\to\infty} a_k = a$.

EXERCISE 1.6.15. Show that if $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are Cauchy sequences in $\mathbb{R}$, then $(a_n + b_n)_{n\in\mathbb{N}}$ and $(a_n \cdot b_n)_{n\in\mathbb{N}}$ are Cauchy sequences in $\mathbb{R}$.

DEFINITION 1.6.16. Let S be a subset of $\mathbb{R}$. Then $x \in \mathbb{R}$ is an *accumulation point* of S if, for all $\varepsilon > 0$, we have $((x - \varepsilon, x + \varepsilon) \setminus \{x\}) \cap S \neq \varnothing$.

REMARK 1.6.17. Thus, x is an accumulation point of S if every interval around x contains points of S other than x. Of course, x does not have to be an element of S in order to be an accumulation point of S.

EXERCISE 1.6.18. Find the accumulation points of the following sets:

(i) $S = (0, 1)$;
(ii) $S = \{(-1)^n + \frac{1}{n} \mid n \in \mathbb{N}\}$;
(iii) $S = \mathbb{Q}$;
(iv) $S = \mathbb{Z}$;
(v) S is the set of rational numbers whose denominators are prime.

LEMMA 1.6.19. *Let S be a subset of $\mathbb{R}$. Then every neighborhood of an accumulation point of S contains infinitely many points of S.*

Proof. Let x be an accumulation point of S. Given $\varepsilon > 0$, there is a point $x_1 \in (x - \varepsilon, x + \varepsilon) \cap S$ such that $x_1 \neq x$. Let $\varepsilon_1 = |x - x_1|$. Then, there is a point $x_2 \in (x - \varepsilon_1, x + \varepsilon_1) \cap S$ such that $x_2 \neq x$. Iterating this procedure, we get an infinite set of elements in S which is contained in $(x - \varepsilon, x + \varepsilon)$.

Now here is a big-time theorem!

THEOREM 1.6.20 **(Bolzano-Weierstrass)**. *Let S be a bounded, infinite subset of $\mathbb{R}$. Then S has an accumulation point.*

Proof. Pick an infinite sequence $(a_k)_{k\in\mathbb{N}}$ of distinct elements of S. Then, by Lemma 1.6.10, $(a_k)_{k\in\mathbb{N}}$ has a convergent subsequence, $(b_j)_{j\in\mathbb{N}}$. If $\lim_{j\to\infty} b_j = b$, then b is an accumulation point of S.

EXERCISE 1.6.21.

(i) Find an infinite subset of $\mathbb{R}$ which does not have an accumulation point.
(ii) Find a bounded subset of $\mathbb{R}$ which does not have an accumulation point.

DEFINITION 1.6.22. Let S be a subset of $\mathbb{R}$. We say that S is an *open set* in $\mathbb{R}$ if, for each point $x \in S$, there is an $\varepsilon > 0$ (depending on x) such that $(x - \varepsilon, x + \varepsilon) \subseteq S$.

DEFINITION 1.6.23. Let $S \subseteq \mathbb{R}$. We say S is a *closed set* in $\mathbb{R}$ if the complement of S is an open set in $\mathbb{R}$.

Note that the empty set and $\mathbb{R}$ are both open and closed subsets of $\mathbb{R}$.

EXERCISE 1.6.24. (i) Show that $\varnothing$ and $\mathbb{R}$ are the only subsets of $\mathbb{R}$ which are both open and closed in $\mathbb{R}$.
(ii) Show that every nonempty open set in $\mathbb{R}$ can be written as a countable union of pairwise disjoint open intervals.
(iii) Show that an arbitrary union of open sets in $\mathbb{R}$ is open in $\mathbb{R}$.
(iv) Show that a finite intersection of open sets in $\mathbb{R}$ is open in $\mathbb{R}$.
(v) Show, by example, that an infinite intersection of open sets is not necessarily open.
(vi) Show that an arbitrary intersection of closed sets in $\mathbb{R}$ is a closed set in $\mathbb{R}$.
(vii) Show that a finite union of closed sets in $\mathbb{R}$ is a closed set in $\mathbb{R}$.
(viii) Show, by example, that an infinite union of closed sets in $\mathbb{R}$ is not necessarily a closed set in $\mathbb{R}$.

EXERCISE 1.6.25. Show that a subset of $\mathbb{R}$ is closed iff it contains all its accumulation points.

EXERCISE 1.6.26. In this exercise, we define the *Cantor set*. This is a subset of the closed interval $[0, 1]$ constructed as follows. First, remove the open interval $(1/3, 2/3)$ from $[0, 1]$. Next, remove the open intervals $(1/9, 2/9)$ and $(7/9, 8/9)$. At each step, remove the middle third of the remaining closed intervals. Repeating this process a countable number of times, we are left with a subset of the closed interval $[0, 1]$ called the *Cantor set*. Show that:

(i) the Cantor set is closed;
(ii) the Cantor set is uncountable;
(iii) the Cantor set consists of all numbers in the closed interval $[0, 1]$ whose ternary expansion consists of only 0's and 2's and may end in infinitely many 2's;
(iv) every point of the Cantor set is an accumulation point of the Cantor set;
(v) the set $[0, 1] \setminus \{\text{Cantor set}\}$ is a dense subset of $[0, 1]$.

The next theorem, the Heine-Borel theorem for $\mathbb{R}$, is the second of the two basic topological theorems for the real numbers; the other is the Bolzano-Weierstrass theorem. We shall see more details about these two theorems in Chapter 2.

THEOREM 1.6.27 **(Heine-Borel)**. *Let S be a closed and bounded subset of $\mathbb{R}$. Given a collection $\{U_i\}_{i\in I}$ of open sets such that $S \subseteq \bigcup_{i\in I} U_i$, there exists a finite subcollection $U_1, \ldots, U_n$ of $\{U_i\}_{i\in I}$ such that $S \subseteq U_1 \cup \cdots \cup U_n$.*

Proof. Suppose that S is a nonempty, closed, bounded subset of $\mathbb{R}$. If $a = \text{glb } S$ and $b = \text{lub } S$, then, since S is closed, a and b are in S, and $S \subseteq [a, b]$. Let $\{U_i\}$ be a collection of open sets such that $S \subseteq \bigcup U_i$. By adjoining the complement of S (if necessary), we obtain a collection $\mathcal{U}$ of open sets whose union contains $[a, b]$.

Now let $B = \{x \in [a, b] \mid [a, x]$ is covered by a finite number of open sets in $\mathcal{U}\}$. Then B is nonempty since $a \in B$, and B is bounded above by b. Let $c = \text{lub } B$. If $c = b$, we are done. If $c < b$, then there exists y such that $c < y < b$ and $[c, y]$ is in the same open set that contains c. Thus $[a, y]$ is covered by the same collection of open sets from $\mathcal{U}$ that covers $[a, c]$. This is a contradiction, and hence b must equal c. Thus $[a, b]$ is covered by a finite number of open sets from $\mathcal{U}$, and by throwing away the complement of S (if necessary), S is covered by a finite number of open sets from the original collection.

DEFINITION 1.6.28. Let A be a subset of $\mathbb{R}$. An *open cover of A* is a collection of open sets $\{U_i\}_{i\in I}$ such that $A \subseteq \bigcup_{i\in I} U_i$.

DEFINITION 1.6.29. Let A be a subset of $\mathbb{R}$. We say that A is a *compact set* if every open covering of A has a finite subcovering. That is, if $\{U_i\}_{i\in I}$ is an open covering of A, then there is a finite subcollection $U_1, U_2, \ldots, U_n$ of the collection $\{U_i\}_{i\in I}$ so that $A \subseteq U_1 \cup U_2 \cup \cdots \cup U_n$.

DEFINITION 1.6.30. A subset A of $\mathbb{R}$ is *sequentially compact* if every infinite sequence in A has a subsequence that converges to an element of A.

EXERCISE 1.6.31. Show that a subset of $\mathbb{R}$ is compact if and only if it is closed and bounded.

The result in the following exercise can be proved easily using the Bolzano-Weierstrass and Heine-Borel theorems in $\mathbb{R}$. We will see in Chapter 2 that the same theorem is true in metric spaces. In the next section, we give an indication of how this works in $\mathbb{C}$.

EXERCISE 1.6.32. A subset of $\mathbb{R}$ is compact if and only if it is sequentially compact.

1.7. Automorphisms of Fields

For any field F, we can consider the following problem: given $f : F \longrightarrow F$ such that $f(x + y) = f(x) + f(y)$ and $f(xy) = f(x)f(y)$ for all $x, y \in F$, what can you say about f? Well, if $f(x) = 0$ for all $x \in F$, then it clearly has these properties, but it is not of much use. So, let us assume that there exists an $a \in F$ such that $f(a) \neq 0$.

EXERCISE 1.7.1. Under this assumption,

(i) show that $f(1) = 1$, and
(ii) show that f is an injection.
(iii) Must such an f be a surjection?

Consider f satisfying only the additive property above in the case when $F = \mathbb{Q}$. We have

$$f(x) = f(x + 0) = f(x) + f(0),$$

so $f(0) = 0$. Next we have, for $n \in \mathbb{N}$, $f(n) = nf(1)$ and $f(m/n) = (m/n)f(1)$ for any positive rational number m/n. While we are not assuming the multiplicative property here, to avoid the trivial, we assume that there exists an $x \in \mathbb{Q}$, so that $f(x) \neq 0$. From this, it follows that $f(1) \neq 0$. Also, for any positive rational number r, we have $f(-r) = -f(r)$. Thus, $f(r) = rf(1)$ for all $r \in \mathbb{Q}$.

Let us see what happens for $\mathbb{R}$. We assume that $f(1) \neq 0$. We have $f(r) = rf(1)$ for all $r \in \mathbb{Q}$. However, as we point out after the definition of Hamel basis in Appendix B, things go completely awry unless we impose further properties. So, assume that f preserves multiplication, that is $f(xy) = f(x)f(y)$ for all $x, y \in \mathbb{R}$. Then, it follows from the above exercise that $f(1) = 1$, and $f(a^{-1}) = f(a)^{-1}$ if $a \neq 0$. The next thing to note here is that if $a \in \mathbb{R}$ and $a \neq 0$, then $a^2 > 0$ and $f(a^2) = (f(a))^2$, so $f(a^2) > 0$. Since all positive real numbers have unique positive square roots, we can conclude that if $c > 0$, then $f(c) > 0$. Thus, if $a < b$, then $f(a) < f(b)$ since $b - a > 0$. Now take any real number c. If $c \in \mathbb{Q}$, then $f(c) = c$. If $c \notin \mathbb{Q}$ and $f(c) \neq c$, then there are two possibilities. If $c < f(c)$, choose a rational number r so that $c < r < f(c)$. Then $f(c) < f(r) = r$, which is a contradiction. If $f(c) < c$, we run into the same problem. So we conclude that f is the identity.

DEFINITION 1.7.2. Let F be a field. An *automorphism of* F is a bijection, $f : F \longrightarrow F$, such that

(a) $f(x + y) = f(x) + f(y)$ for all $x, y \in F$,
(b) $f(xy) = f(x)f(y)$ for all $x, y \in F$.

EXERCISE 1.7.3. If F is a field, show that the automorphisms of F form a group under composition of functions. This group is called *the automorphism group of the field* F and is denoted by $\mathrm{Aut}\,(F)$.

THEOREM 1.7.4. *The groups* $\mathrm{Aut}\,(\mathbb{Q})$ *and* $\mathrm{Aut}\,(\mathbb{R})$ *consist only of the identity map.*

EXERCISE 1.7.5. Find a field F such that $\mathrm{Aut}\,(F) \neq \{1\}$.

EXERCISE 1.7.6.

(i) Let F be a field and let f be an element of $\mathrm{Aut}\,(F)$. Define $H_f = \{x \in F \mid f(x) = x\}$. Show that H_f is a subfield of F.
(ii) Suppose that F is a field and that $\mathbb{Q}$ is a subfield of F. If $f \in \mathrm{Aut}\,(F)$, show that $\mathbb{Q}$ is a subfield of H_f.

EXERCISE 1.7.7.

(i) Find $\mathrm{Aut}\,(\mathbb{F}_p)$ where p is a prime and $\mathbb{Z}_p$ is the finite field with p elements.

(ii) Let $F = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. Show that F is a field and find $\mathrm{Aut}\,(F)$. This is the beginning of the subject called Galois theory, in which one of the goals is to determine $\mathrm{Aut}\,(F)$ when F is a so-called "algebraic extension" of $\mathbb{Q}$.

1.8. Complex Numbers

To start this section, we give a somewhat inexact definition of complex numbers. This is often used as a definition of the complex numbers, but it does contain some ambiguity, which we will rectify immediately.

DEFINITION 1.8.1 **(Rural definition)**. The set of *complex numbers*, $\mathbb{C}$, is the collection of expressions of the form $z = a + bi$ where $a, b \in \mathbb{R}$ and i is a symbol which satisfies $i^2 = -1$. If $z = a + bi$ and $w = c + di$ are in $\mathbb{C}$, then we define $z + w = (a + c) + (b + d)i$ and $zw = (ac - bd) + (bc + ad)i$.

Actually, one can go a long way with this definition if the symbol i with the property that $i^2 = -1$ does not cause insomnia. To be more precise, we consider the Cartesian product $\mathbb{R}\times\mathbb{R}$ with addition defined by $(a, b)+(c, d) = (a + c, b + d)$ and multiplication defined by $(a, b)(c, d) = (ac - bd, bc + ad)$.

EXERCISE 1.8.2. Show that $\mathbb{R} \times \mathbb{R}$ with addition and multiplication as defined above is a field with $(0, 0)$ as the additive identity, $(1, 0)$ as the multiplicative identity, $-(a, b) = (-a, -b)$, and $(a, b)^{-1} = (a/(a^2+b^2), -b/(a^2 + b^2))$ if $(a, b) \neq (0, 0)$.

So $\mathbb{R} \times \mathbb{R}$ with these operations forms a field which we denote by $\mathbb{C}$ and call *the field of complex numbers*. Note that $\mathbb{R}$ is isomorphic to the subfield of $\mathbb{C}$ given by $\{(a, 0) \mid a \in \mathbb{R}\}$. If we set $i = (0, 1)$, then $i^2 = (-1, 0)$. Finally, to fix things up really nice, we write $(a, b) = (a, 0)+(b, 0)(0, 1)$, or, returning to our original rural definition, $(a, b) = a + bi$.

The first observation to make is that $\mathbb{C}$ cannot be made into an ordered field. That is, it cannot satisfy the order axioms given in Appendix A. This is immediate because in any ordered field, if $a \neq 0$, then $a^2 > 0$. This would imply that $i^2 = -1 > 0$, but $1^2 = 1 > 0$, and this is a contradiction.

DEFINITION 1.8.3. If $z = a + bi$, we call a the *real part* of z and b the *imaginary part* of z. We write $a = \mathrm{Re}\ z$ and $b = \mathrm{Im}\ z$. The complex number z is called *pure imaginary* if $a = \mathrm{Re}\ z = 0$.

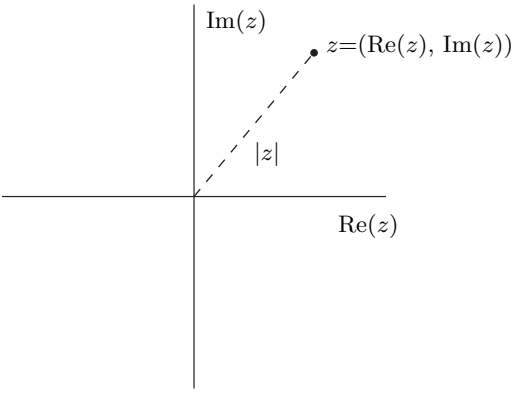
DEFINITION 1.8.4. If $z = a + bi$, the *complex conjugate* of z, denoted $\bar{z}$, is $\bar{z} = a - bi$. The *absolute value* of z is

$$|z| = (z\bar{z})^{\frac{1}{2}} = (a^2 + b^2)^{\frac{1}{2}},$$

where, of course, we mean the nonnegative square root in $\mathbb{R}$.

If z and w are complex numbers, then $|z|, |w| \in \mathbb{R}$ and hence it makes sense to say that $|z| < |w|$. However, it makes no sense to say that $z < w$.

EXERCISE 1.8.5. Show that if we identify $z = a + bi$ with the point $(a, b) \in \mathbb{R}^2$, then the absolute value of z is equal to the distance of the point (a, b) from $(0, 0)$.



EXERCISE 1.8.6. Show that the absolute value on $\mathbb{C}$ satisfies all the properties of the absolute value on $\mathbb{R}$.

(i) For any $z \in \mathbb{C}$, we have $|z| \geq 0$, and $|z| = 0$ iff $z = 0$.
(ii) For any $z, w \in \mathbb{C}$, we have $|zw| = |z||w|$.
(iii) For any $z, w \in \mathbb{C}$, we have $|z + w| \leq |z| + |w|$ (triangle inequality).

EXERCISE 1.8.7. Show that the field of complex numbers is not isomorphic to the field of real numbers.

1.9. Convergence in $\mathbb{C}$

Now that we have an absolute value on $\mathbb{C}$, we can define the notions of Cauchy sequence and convergent sequence in $\mathbb{C}$.

DEFINITION 1.9.1. A sequence $(z_k)_{k\in\mathbb{N}}$ of complex numbers is *convergent* if there exists an element $z \in \mathbb{C}$ such that the sequence satisfies the following property: given any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $k \geq N$ implies that $|z_k - z| < \varepsilon$. We say that $(z_k)_{k\in\mathbb{N}}$ *converges to* z, and z is called the *limit* of the sequence $(z_k)_{k\in\mathbb{N}}$. Symbolically, we write

$$\lim_{k\to\infty} z_k = z.$$

We will often say that a sequence of complex numbers is convergent without specific reference to the limit z. Note that N depends on ε. As usual, the limit of a convergent sequence is unique.

DEFINITION 1.9.2. Let r be a positive real number, and let $z_0 \in \mathbb{C}$. The *open ball* of radius r with center at z_0 is

$$B_r(z_0) = \{z \in \mathbb{C} \mid |z - z_0| < r\}. \tag{1.1}$$

The *closed ball* of radius r with center z_0 is

$$\bar{B}_r(z_0) = \{z \in \mathbb{C} \mid |z - z_0| \leq r\}. \tag{1.2}$$

The open balls and closed balls in $\mathbb{C}$ are the analogs of open and closed intervals in $\mathbb{R}$. We can define open and closed sets in $\mathbb{C}$ in a fashion similar to the definitions in $\mathbb{R}$.

DEFINITION 1.9.3. Let S be a subset of $\mathbb{C}$. We say that S is an *open set* in $\mathbb{C}$ if, for each point $z \in S$, there is an $\varepsilon > 0$ (depending on z) such that $B_\varepsilon(z) \subseteq S$.

DEFINITION 1.9.4. Let S be a subset of $\mathbb{C}$. We say that S is a *closed set* in $\mathbb{C}$ if the complement of S is an open set in $\mathbb{C}$.

Note that the empty set and $\mathbb{C}$ are both open and closed subsets of $\mathbb{C}$.

EXERCISE 1.9.5. (i) Show that $\varnothing$ and $\mathbb{C}$ are the only subsets of $\mathbb{C}$ which are both open and closed in $\mathbb{C}$.
(ii) Show that every open set in $\mathbb{C}$ can be written as a countable union of open balls.
(iii) Show, by example, that there are open sets in $\mathbb{C}$ for which the open balls in (ii) cannot be made pairwise disjoint.
(iv) Show that an arbitrary union of open sets in $\mathbb{C}$ is an open set in $\mathbb{C}$.
(v) Show that a finite intersection of open sets in $\mathbb{C}$ is an open set in $\mathbb{C}$.
(vi) Show, by example, that an infinite intersection of open sets in $\mathbb{C}$ need not be an open set in $\mathbb{C}$.
(vii) Show that an arbitrary intersection of closed sets in $\mathbb{C}$ is a closed set in $\mathbb{C}$.
(viii) Show that a finite union of closed sets in $\mathbb{C}$ is a closed set in $\mathbb{C}$.
(ix) Show, by example, that an infinite union of closed sets in $\mathbb{C}$ is not necessarily a closed set in $\mathbb{C}$.

EXERCISE 1.9.6. Consider the collection of open balls $\{B_r(z)\}$ in $\mathbb{C}$ where $r \in \mathbb{Q}$ and where Re z and Im $z \in \mathbb{Q}$. Show that any open set in $\mathbb{C}$ can be written as a finite or countable union from this collection of sets.

DEFINITION 1.9.7. Let $A \subseteq \mathbb{C}$. The set A is *bounded* if there exists $r > 0$ such that $A \subseteq B_r(0)$.

EXERCISE 1.9.8. Define the notion of a bounded sequence in $\mathbb{C}$.

DEFINITION 1.9.9 **(**See Definition 1.6.12). A sequence $(z_k)_{k\in\mathbb{N}}$ in $\mathbb{C}$ is a *Cauchy sequence* if, given any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $n, m \geq N$ implies $|z_n - z_m| < \varepsilon$.

EXERCISE 1.9.10. Prove that every Cauchy sequence in $\mathbb{C}$ is bounded.

THEOREM 1.9.11 **(Cauchy criterion)**. *A sequence $(z_k)_{k\in\mathbb{N}}$ of complex numbers is convergent if and only if it is a Cauchy sequence.*

Proof. The first half of the proof is identical to the proof of Theorem 1.6.14.

Suppose now that $(z_k)_{k\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{C}$. Let $z_k = a_k + b_k i$, where $a_k, b_k \in \mathbb{R}$. Then $|z_m - z_n|^2 = (a_m - a_n)^2 + (b_m - b_n)^2$. It follows immediately that $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ are Cauchy sequences in $\mathbb{R}$. If $\lim_{k\to\infty} a_k = a$ and $\lim_{k\to\infty} b_k = b$, then $\lim_{k\to\infty} z_k = z$ where $z = a + bi$.

EXERCISE 1.9.12. Show that every bounded sequence in $\mathbb{C}$ has a convergent subsequence.

DEFINITION 1.9.13. Let S be a subset of $\mathbb{C}$. Then z is an *accumulation point* of S if, for all $\varepsilon > 0$, we have $(B_\varepsilon(z) \setminus \{z\}) \cap S \neq \varnothing$.

REMARK 1.9.14. Thus, z is an accumulation point of S if every open ball around z contains points of S other than z. Of course, z does not have to be an element of S in order to be an accumulation point of S.

EXERCISE 1.9.15. Find the accumulation points of the following sets:

(i) $S = \{z \in \mathbb{C} \mid |z| = 1\}$ (this is the *unit circle in* $\mathbb{C}$);
(ii) $S = \{z \in \mathbb{C} \mid \text{Re } z > \text{Im } z\}$;
(iii) $S = \{a + bi \mid a, b \in \mathbb{Q}\}$;
(iv) $S = \{a + bi \mid a, b \in \mathbb{Z}\}$.

EXERCISE 1.9.16.

(i) Let S be a subset of $\mathbb{C}$. Show that every neighborhood of an accumulation point of S contains infinitely many points of S.
(ii) (Bolzano-Weierstrass theorem for $\mathbb{C}$) Prove that any bounded infinite set in $\mathbb{C}$ has an accumulation point in $\mathbb{C}$.

THEOREM 1.9.17 **(Heine-Borel)**. *Let S be a closed and bounded subset of $\mathbb{C}$. Given a collection $\{U_i\}_{i\in I}$ of open sets such that $S \subseteq \bigcup_{i\in I} U_i$, there exists a finite subcollection $U_1, \ldots, U_n$ of $\{U_i\}_{i\in I}$ such that $S \subseteq U_1 \cup \cdots \cup U_n$.*

Proof. For the purposes of this proof, we treat $\mathbb{C}$ as $\mathbb{R}^2$. We prove it for $S = [a, b] \times [c, d]$ where $a, b, c, d \in \mathbb{R}$ and $a < b$ and $c < d$ and leave the general case as an exercise. Take a point $x_0 \in [a, b]$ and consider the set $\{x_0\} \times [c, d]$. We take an open set $N \subseteq \mathbb{C}$ containing $\{x_0\} \times [c, d]$. We claim that there exists an open interval I around x_0 such that $I \times [c, d] \subseteq N$. For each point in $(x_0, y) \in \{x_0\} \times [c, d]$, choose $r_y > 0$ such that the open square $(x_0 - r_y, x_0 + r_y) \times (y - r_y, y + r_y) \subseteq N$. By intersecting these squares with $\{x_0\} \times \mathbb{R}$ and projecting on the second coordinate, we get a collection of open intervals of the form $\{x_0\} \times (y - r_y, y + r_y)$ that cover $\{x_0\} \times [c, d]$. By the Heine-Borel theorem in $\mathbb{R}$, there exists a finite subcollection of these open intervals that covers the interval $\{x_0\} \times [c, d]$. Hence the corresponding collection of open squares also covers $\{x_0\} \times [c, d]$. Let r be the minimum of the r_y from this finite collection. Then $I = (x_0 - r, x_0 + r)$ is the interval we sought.

Now let $\{U_j\}_{j\in J}$ be an open covering of S. For each $x \in [a, b]$, the collection $\{U_j\}_{j\in J}$ covers $\{x\} \times [c, d]$. As we did above, we choose a finite

subcollection $U_1, \ldots, U_n$ that covers $\{x\} \times [c, d]$. The open set $N_x = U_1 \cup \cdots \cup U_n$ contains a set of the form $I_x \times [c, d]$ by the preceding discussion, where I_x is an open interval containing x. The collection $\{I_x\}_{x \in [a,b]}$ covers $[a, b]$, and hence by the Heine-Borel theorem for $\mathbb{R}$, there exists a finite subcollection $I_{x_1}, \ldots, I_{x_m}$ that covers $[a, b]$. We take our finite subcollection of the original open cover $\{U_j\}_{j \in J}$ to be $\{U \mid \text{for some } x_i \text{ the set } U \text{ is one of the elements in the union that defines } N_{x_i}\}$.

EXERCISE 1.9.18. Show that a subset of $\mathbb{C}$ is closed iff it contains all its accumulation points.

EXERCISE 1.9.19. Define the notion of sequentially compact for a subset of $\mathbb{C}$, and show that a subset of $\mathbb{C}$ is sequentially compact if and only if it is closed and bounded.

DEFINITION 1.9.20. If $z = x + iy \in \mathbb{C}$, $z \neq 0$, and $r = |z|$, then the *polar form* of z is $z = r(\cos\theta + i \sin\theta)$ where θ is the unique solution to the equations

$$x = r\cos\theta,$$
$$y = r\sin\theta$$

in the interval $[0, 2\pi)$. The angle θ is called *the principal branch of the argument of* z and is denoted $\mathrm{Arg}(z)$. For z as above, we often write $z = re^{i\theta}$ where $e^{i\theta}$ is defined to be $\cos\theta + i\sin\theta$. In fact, this is the value of the complex exponential function $f(z) = e^z$ when $z = i\theta$.

EXERCISE 1.9.21. Suppose that $n \in \mathbb{N}$. Prove that if $z = e^{\frac{2\pi ik}{n}}$, for $k \in \mathbb{Z}$ and $0 \leq k \leq n - 1$, then $z^n = 1$. Such a z is called an *n-th root of unity*. Note that these are all distinct.

The n-th roots of unity form a cyclic group of order n under multiplication. An n-th root of unity is *primitive* if it is a generator of this group.

EXERCISE 1.9.22. Show that the primitive n-th roots of unity are of the form $e^{2\pi ik/n}$ where k and n are relatively prime.

PROPOSITION 1.9.23. *If $n > 1$, the sum of the n distinct n-th roots of unity is 0.*

Proof. For any $z \in \mathbb{C}$,

$$1 - z^n = (1 - z)(1 + z + z^2 + \cdots + z^{n-1}).$$

Now let z be a primitive n-th root of unity.

EXERCISE 1.9.24. Suppose z is a nonzero complex number, and write $z = re^{i\theta}$. Show that z has exactly n distinct complex n-th roots given by $r^{1/n}e^{i(2\pi k+\theta)/n}$ for $0 \leq k \leq n - 1$.

We now consider the ring of polynomials $\mathbb{C}[z]$ over $\mathbb{C}$.

EXERCISE 1.9.25. Show that $\mathbb{C}[z]$ is an integral domain. Determine the elements in this domain which have multiplicative inverses.

DEFINITION 1.9.26. Let F be a field. We say that F is *algebraically closed* if every nonconstant polynomial in $F[x]$ has a root in F. That is, F is algebraically closed if, for every nonconstant $p(x) \in F[x]$, there is an element $r \in F$ such that $p(r) = 0$.

The most important example of an algebraically closed field is the complex numbers. There are a semi-infinite number of proofs of this theorem. We will present one of these as a project in Section 2.7.2 using the properties of continuous functions developed in Chapter 2.

EXERCISE 1.9.27. Let F be a field and suppose that $p(x) \in F[x]$. Show that r is a root of $p(x)$ if and only if $(x - r)$ is a factor of $p(x)$. That is, we can write $p(x) = (x - r)q(x)$ for some $q(x) \in F[x]$.

DEFINITION 1.9.28. Let $\mathbb{A}$ be the collection of all complex roots of polynomials in $\mathbb{Z}[x]$. The set $\mathbb{A}$ is called *the set of algebraic numbers* in $\mathbb{C}$. The set $\mathbb{A}_{\mathbb{R}} = \mathbb{A} \cap \mathbb{R}$ is called *the set of real algebraic numbers*. A real number which is not a real algebraic number is called *transcendental*.

EXAMPLE 1.9.29. Among the more famous algebraic numbers are i and $-i$. For real algebraic numbers, the most famous one is probably $\sqrt{2}$. The most famous transcendental numbers are π and e.

EXERCISE 1.9.30. Show that $\mathbb{A}$ and $\mathbb{A}_{\mathbb{R}}$ are fields.

EXERCISE 1.9.31. Show that the field $\mathbb{A}$ of algebraic numbers is countable.

REMARK 1.9.32. It follows from the exercise above that the field $\mathbb{A}_{\mathbb{R}}$ of real algebraic numbers is countable and hence the set of transcendental numbers is uncountable.

1.10. Independent Projects

1.10.1. Another Construction of $\mathbb{R}$.

DEFINITION 1.10.1. A subset α of $\mathbb{Q}$ is said to be a *cut* (or a *Dedekind cut*) if it satisfies the following:

(a) the set $\alpha \neq \varnothing$ and $\alpha \neq \mathbb{Q}$;
(b) if $r \in \alpha$ and if $s \in \mathbb{Q}$ satisfies $s < r$, then $s \in \alpha$;
(c) if $r \in \alpha$, then there exists $s \in \mathbb{Q}$ with $s > r$ and $s \in \alpha$.

Let R denote the collection of all cuts.

DEFINITION 1.10.2. For $\alpha, \beta \in R$, we define $\alpha + \beta = \{r + s \mid r \in \alpha \text{ and } s \in \beta\}$. Let $\mathbf{0} = \{r \in \mathbb{Q} \mid r < 0\}$.

EXERCISE 1.10.3. If α and β are cuts, show that $\alpha+\beta$ is a cut, and also show that $\mathbf{0}$ is a cut.

EXERCISE 1.10.4. Show that with this addition $(R,+)$ is an abelian group with $\mathbf{0}$ as the identity element.

We now define an order on R.

DEFINITION 1.10.5. If $\alpha, \beta \in R$, we say that $\alpha < \beta$ if α is a proper subset of β.

EXERCISE 1.10.6. Show that the relation $<$ satisfies the following properties:

(i) (trichotomy) if $\alpha, \beta \in R$, then one and only one of the following holds: $\alpha < \beta$, $\alpha = \beta$, or $\beta < \alpha$;
(ii) (transitivity) if $\alpha, \beta, \gamma \in R$ with $\alpha < \beta$ and $\beta < \gamma$, then $\alpha < \gamma$;
(iii) (additivity) if $\alpha, \beta, \gamma \in R$ with $\alpha < \beta$, then $\alpha + \gamma < \beta + \gamma$.

It is now possible to define the notions of bounded above, bounded below, bounded, upper bound, least upper bound, lower bound, and greatest lower bound in R just as we did earlier in this chapter.

EXERCISE 1.10.7. Show that the least upper bound property holds in R, that is, if A is a nonempty subset of R which is bounded above, then A has a least upper bound in R.

Next, we must define multiplication in R.

DEFINITION 1.10.8. If $\alpha, \beta \in R$ with $\alpha, \beta > 0$, then

$$\alpha\beta = \{p \in \mathbb{Q} \mid \text{there are positive elements } r \in \alpha \text{ and } s \in \beta \text{ so that } p \leq rs\}.$$

The next step is multiplication by $\mathbf{0}$, which is exactly as it should be, namely for any $\alpha \in R$, we define $\alpha\mathbf{0} = \mathbf{0}$.

EXERCISE 1.10.9. If $\alpha < 0$ or $\beta < 0$ or both, replace any negative element by its additive inverse and use the multiplication of positive elements to define multiplication accordingly. For example, if $\alpha < 0$ and $\beta > 0$, $\alpha\beta = -[(-\alpha)(\beta)]$. Show that R with addition, multiplication, and order as defined above is an ordered field.

EXERCISE 1.10.10. Put it all together and show that R is an Archimedean ordered field in which the least upper bound property holds.

1.10.2. Infinite Series.

An important topic in analysis is the study of infinite series. This theory will be used in the remaining chapters of this book.

We assume that the reader has had at least an elementary introduction to infinite series and their convergence properties. In fact, the theory of infinite series actually reduces to the convergence of sequences, which we have covered thoroughly in this chapter. An infinite series is expressed as

a sum of an infinite number of elements from some place or other. These elements could be numbers, functions, or what have you, so we begin with one-sided series of numbers.

An infinite series is an expression of the form $\sum_{n=1}^{\infty} a_n$, where the elements a_n come from a number system in which addition makes sense. So that we don't wander around aimlessly, let's fix our number system to be the complex numbers, that is, $a_n \in \mathbb{C}$, with the possibility of restricting ourselves to the real numbers or even the rational numbers. In the definition we have chosen to use the natural numbers as the index set, but in considering infinite series, we could start the summation with any integer and write $\sum_{n=n_0}^{\infty} a_n$. Later, we will also consider two-sided series where the index set is the integers and we write $\sum_{-\infty}^{\infty} a_n$. If these expressions are going to have any meaning at all, we must look at the partial sums.

DEFINITION 1.10.11. If $\sum_{n=1}^{\infty} a_n$ is an infinite series of complex numbers, the *N-th partial sum* of the series is $S_N = \sum_{n=1}^{N} a_n$.

EXAMPLES 1.10.12.

(i) Let $a_n = 1$ for all n. Then $S_N = N$.
(ii) Let $a_n = 1/n$. Then $S_N = 1 + 1/2 + \cdots + 1/N$.
(iii) Let $a_n = 1/2^n$. Then $S_N = 1 - 1/2^N$.
(iv) Let $a_n = (-1)^{n+1}$. In this case, $S_N = 1$ if N is odd and 0 if N is even.
(v) Fix θ, with $0 < \theta < 2\pi$, and let $a_n = e^{in\theta}/n$. Then $S_N = \sum_{n=1}^{N} e^{in\theta}/n$, which is the best we can do without more information about θ.
(vi) Let $a_n = \sin n\pi/n^2$. In this case, $S_N = \sum_{n=1}^{N} \sin(n\pi)/n^2$.

DEFINITION 1.10.13. Let $\sum_{n=1}^{\infty} a_n$ be an infinite series of complex numbers. The sequence $(S_N)_{N\in\mathbb{N}}$ is called *the sequence of partial sums.* We say that the series $\sum_{n=1}^{\infty} a_n$ *converges* if the sequence of partial sums $(S_N)_{N\in\mathbb{N}}$ converges. If the sequence $(S_N)_{N\in\mathbb{N}}$ does not converge, we say that $\sum_{n=1}^{\infty} a_n$ *diverges.*

Of course, since we are working in $\mathbb{C}$, the series converges if and only if the sequence $(S_N)_{N\in\mathbb{N}}$ is a Cauchy sequence. That is, given $\varepsilon > 0$, there is an $N \in \mathbb{N}$ so that for $n, m > N$ (assuming $n > m$), $|\sum_{k=m+1}^{n} a_n| < \varepsilon$.

EXERCISE 1.10.14. Determine which of the series in Example 1.10.12 converge.

We are faced with two problems. The first is, "How do we tell if a series converges?" The second is, "If a series does converge, how do we find the explicit sum?" There is extensive literature about these two questions, but the fact is that the second question presents many more difficulties than the first. In Chapter 7, the theory of Fourier series will provide some assistance.

The most helpful series in all of this discussion is the geometric series.

DEFINITION 1.10.15. Let z be a complex number. The geometric series defined by z is $\sum_{n=0}^{\infty} z^n$.

EXERCISE 1.10.16.

(i) If $N \in \mathbb{N}$ and $z \neq 1$, show that $S_N = \sum_{n=0}^{N} z^n = \frac{1-z^{N+1}}{1-z}$.
(ii) If $|z| < 1$, show that $\lim_{n\to\infty} z^n = 0$.
(iii) If $|z| > 1$, show that $\lim_{n\to\infty} z^n$ does not exist.

THEOREM 1.10.17. *Consider the geometric series defined by a complex number z. If $|z| < 1$, then the series converges. If $|z| > 1$, then the series diverges.*

Proof. This follows from the exercise above.

EXERCISE 1.10.18.

(i) What can you say if $|z| = 1$?
(ii) Suppose that a series $\sum_{n=1}^{\infty} a_n$ converges. Show that $\lim_{n\to\infty} a_n = 0$.

The property $\lim_{n\to\infty} a_n = 0$ does not ensure that the series $\sum_{n=1}^{\infty} a_n$ converges. The most useful example is given above where $a_n = 1/n$. In this case, $S_1 = 1$, $S_4 > 2$, and it is easy to check that $S_{2^n} > n$ for $n \in \mathbb{N}$. It follows that the series $\sum_{n=1}^{\infty} 1/n$ diverges. The series $S = \sum_{n=1}^{\infty} 1/n$ is often called the *harmonic series*. We have just proved that this series diverges.

EXERCISE 1.10.19.

(i) Let $S_N = \sum_{n=1}^{N} 1/n$. Show that, for $N \geq 2$, S_N is never an integer.
(ii) Show that, by suitably eliminating an infinite number of terms, the remaining subseries can be made to converge to any positive real number.

EXERCISE 1.10.20.

(i) If $s \in \mathbb{R}$ and $s > 1$, show that $\sum_{n=1}^{\infty} 1/n^s$ converges.
(ii) If $s \in \mathbb{R}$ and $s < 1$, show that $\sum_{n=1}^{\infty} 1/n^s$ diverges.
(iii) For which $s \in \mathbb{R}$ does the series $\sum_{p \text{ prime}} 1/p^s$ converge?

DEFINITION 1.10.21. A series $\sum_{n=1}^{\infty} a_n$ of complex numbers *converges absolutely* if the series $\sum_{n=1}^{\infty} |a_n|$ converges.

PROPOSITION 1.10.22. *If $\sum_{n=1}^{\infty} a_n$ converges absolutely, then $\sum_{n=1}^{\infty} a_n$ converges*

Proof. This follows from the fact that $|\sum_{k=m+1}^{n} a_k| \leq \sum_{k=m+1}^{n} |a_k|$.

The converse to Proposition 1.10.22 is false and is shown by the example $\sum_{n=1}^{\infty} (-1)^{n+1}/n$. This series converges since $|\sum_{k=m+1}^{n} (-1)^{k+1}/k| < 1/m$. However, as we have seen above, the series does not converge absolutely.

There are various tests to determine if a series converges. These include the comparison test, the ratio test, and the root test. The comparison test is often very useful, but its use depends on knowing ahead of time a series which converges.

THEOREM 1.10.23 **(Comparison test)**. *Suppose $a_n > 0$ for $n \in \mathbb{N}$ and suppose $\sum_{n=1}^{\infty} a_n$ converges. If $b_n \in \mathbb{C}$ satisfies $|b_n| \leq a_n$ for all n, then the series $\sum_{n=1}^{\infty} b_n$ converges absolutely and hence converges.*

EXERCISE 1.10.24.

(i) Prove the comparison test.
(ii) If the series $\sum_{n=1}^{\infty} a_n$ converges to s and c is any constant, show that the series $\sum_{n=1}^{\infty} ca_n$ converges to cs.
(iii) Suppose that $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are infinite series. Suppose that $a_n > 0$ and $b_n > 0$ for $n \in \mathbb{N}$ and $\lim_{n\to\infty} a_n/b_n = c > 0$. Show that $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges.

The most useful series for comparison is the geometric series defined by a real number r, with $0 < r < 1$.

THEOREM 1.10.25 **(Ratio test)**. *Suppose that $\sum_{n=1}^{\infty} a_n$ is a series of nonzero complex numbers. If $r = \lim_{n\to\infty} |a_{n+1}/a_n|$ exists, then the series converges absolutely if $r < 1$ and the series diverges if $r > 1$.*

Proof. Suppose $\lim_{n\to\infty} |a_{n+1}/a_n| = r < 1$. If ρ satisfies $r < \rho < 1$, then there exists $N \in \mathbb{N}$ such that $|a_{n+1}|/|a_n| < \rho$ for all $n \geq N$. Consequently, $|a_n| \leq |a_N|\rho^{n-N}$ for all $n \geq N$. The result follows from the comparison test.

EXERCISE 1.10.26. Show that if $r > 1$, then the series above diverges, while, if $r = 1$, anything can happen.

Our final test for convergence is called the root test. This can be quite effective when the comparison test and ratio test fail.

THEOREM 1.10.27 **(Root test)**. *Suppose that $\sum_{n=1}^{\infty} a_n$ is a series of complex numbers. Let*

$$r = \limsup_{n\to\infty} |a_n|^{1/n}$$

(consult Chapter 2 for a discussion of lim sup*). If $r < 1$, then the series converges absolutely. If $r > 1$, then the series diverges.*

Proof. Suppose that $\limsup_{n\to\infty} |a_n|^{1/n} = r < 1$. Pick ρ so that $r < \rho < 1$. Then, there exists $N \in \mathbb{N}$ such that $|a_n| \leq \rho^n$ for all $n \geq N$. The convergence of the series now follows from the comparison test.

EXERCISE 1.10.28. Show that if $r > 1$, then the above series diverges, while, if $r = 1$, anything can happen.

EXERCISE 1.10.29. Suppose that the ratio test applies to a series. That is, $\lim_{n\to\infty} |a_{n+1}|/|a_n| = r$. Show that the $\limsup_{n\to\infty} |a_n|^{1/n} = r$.

DEFINITION 1.10.30. Let z_0 be a fixed complex number. A *complex power series* around z_0 is a series of the form $\sum_{n=0}^{\infty} a_n(z - z_0)^n$, where the coefficients $a_n \in \mathbb{C}$. When this series converges, it converges to a function of the complex variable z.

EXERCISE 1.10.31. Show that if the series converges absolutely for a complex number z, then it also converges for any complex number w such that $|w - z_0| \leq |z - z_0|$. That is, the series converges on the disk

$$\{w \in \mathbb{C} \mid |w - z_0| \leq |z - z_0|\}.$$

From this exercise, it follows that a complex power series around z_0 that converges absolutely at any point other then z_0 will have a disk of convergence of the form $\{z \in \mathbb{C} \mid |z - z_0| < r\}$. The supremum of all such r is called the *radius of convergence* of the power series. If the power series converges only at the point $z = z_0$, we say that the series has radius of convergence equal to 0.

To determine the radius of convergence for a complex power series, we use the convergence tests developed above, in particular the root test.

THEOREM 1.10.32. *Suppose that* $\limsup_{n\to\infty} |a_n|^{1/n} = r$. *If* $0 < r < \infty$, *then the power series* $\sum_{n=0}^{\infty} a_n(z - z_0)^n$ *has radius of convergence* $1/r$. *If* $r = 0$, *then the radius of convergence is infinity. If* $r = \infty$, *then the radius of convergence is* 0.

EXAMPLES 1.10.33.

(i) Consider the series $\sum_{n=0}^{\infty} n(z - z_0)^n$. Then $\lim_{n\to\infty} n^{1/n} = 1$, and the power series converges absolutely for $|z - z_0| < 1$, that is, the radius of convergence is 1.
(ii) Consider the series $\sum_{n=1}^{\infty} n^n(z - z_0)^n$. Then $\lim_{n\to\infty}(n^n)^{1/n} = \infty$, so the radius of convergence is 0 and the series converges only for $z = z_0$.

EXERCISE 1.10.34. Determine the radius of convergence of the following power series:

(i) $\sum_{n=1}^{\infty} \frac{z^n}{n!}$;
(ii) $\sum_{n=2}^{\infty} \frac{z^n}{\ln(n)}$;
(iii) $\sum_{n=1}^{\infty} \frac{n^n}{n!} z^n$.

1.10.3. Decimal Expansions of Real Numbers.

In Appendix A, we used a decimal representation of the real numbers to show that the real numbers between 0 and 1 form an uncountable set. In this project, we actually prove that every real number between 0 and 1 has a decimal expansion which is unique with the condition that no expansion can terminate in all 9's. In addition, we discuss the fact that rational numbers have decimal expansions of three different types. The first is terminating decimals, the second is rational numbers whose denominators are relatively prime to 10, and the third is a combination of the first two.

Since we know that every real number lies between two consecutive integers, we start with a real number x so that $0 < x < 1$. Let $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Assume first that x is irrational. The construction proceeds as follows. Let a_1 be the largest element of S which is less than $10x$. Then $0 < x - a_1/10 < 1/10$. Let a_2 be the largest integer in S less than $100x - 10a_1$. Proceeding as before, we get $0 < x - a_1/10 - a_2/10^2 < 1/10^2$. Continuing this process, we obtain a monotonic increasing sequence

$$S_n = a_1/10 + a_2/10^2 + \cdots + a_n/10^n,$$

where $a_j \in S$ and $0 < x - S_n < 1/10^n$. We conclude that S_n converges to x and we get

$$x = a_1/10 + a_2/10^2 + \cdots + a_n/10^n + \cdots . \tag{1.3}$$

EXERCISE 1.10.35. Let x be a irrational number between 0 and 1. Show that there is only one way to express x in the form (1.3).

We now turn to rational numbers between 0 and 1. We can apply the above procedure to rational numbers with the possibility of equality in any of the inequalities above. Suppose that x has a *terminating decimal expansion.* That is, suppose there exists N so that $a_n = 0$ for all $n > N$ and $a_N \neq 0$. Then we can write $x = a_1/10 + a_2/10^2 + \cdots + a_N/10^N$.

EXERCISE 1.10.36.

(i) Show that if r is a rational number in $(0, 1)$, then the decimal expansion of r terminates if and only if the denominator of r has the form $2^a 5^b$ where a and b are nonnegative integers and are not both zero.
(ii) With r as above, show that the last nonzero digit of r is in the m-th place where $m = \max(a, b)$.

Note that rational numbers with terminating decimal expansions are the only real numbers between 0 and 1 for which equality can occur in the procedure above.

Next consider a rational number $r = p/q$ in $(0, 1)$ for which q is relatively prime to 10. From Euler's theorem (see the project in Section A.10.1), q divides $10^{\phi(q)} - 1$. Let n be the smallest natural number so that q divides $10^n - 1$. Then $(p/q)(10^n - 1)$ is an integer which we denote by m. That is,

$$m = \frac{p}{q}(10^n - 1) \quad \text{or} \quad \frac{p}{q} = \frac{m}{10^n - 1}.$$

We can now write

$$\begin{aligned} \frac{p}{q} &= \frac{m}{10^n - 1} = \frac{m}{10^n}(1 - 10^{-n})^{-1} = \frac{m}{10^n}(1 + 10^{-n} + 10^{-2n} + \cdots) \\ &= m/10^n + m/10^{2n} + \cdots . \end{aligned}$$

As $0 < p/q < 1$, we have $m < 10^n$. Thus the right-hand side of the equation above gives us a periodic decimal expansion of p/q whose period has length n.

EXERCISE 1.10.37. Let p/q be a rational number between 0 and 1. If q and 10 are relatively prime, show that p/q has a unique periodic decimal expansion with the length of the period equal to the order of 10 mod q.

We now present the remaining case as an exercise.

EXERCISE 1.10.38. Let p/q be a rational number in $(0, 1)$ with $q = 2^a 5^b r$ where r is relatively prime to 10. Let $k = \max(a, b)$ and let n be the smallest positive integer such that r divides $10^n - 1$. Show that, after k digits, the decimal expansion of p/q is periodic of length n.

EXERCISE 1.10.39. Can any of the above decimal expansions terminate in all 9's?

Chapter 2

Metric and Euclidean Spaces

> ...la notion d'espace métrique fut introduite en 1906 par M. Fréchet, et dévelopée quelques années plus tard par F. Hausdorff dans sa Mengenlehre. Elle acquit une grande importance après 1920, d'une part à la suite des travaus fondamentaux de S. Banach et de son école sur les espaces normés et leurs applications à l'Analyse fonctionnelle, de l'autre en raison de l'intérêt que présente la notion de valeur absolue en Arithmétique et en Géométrie algébrique (où notamment la complétion par rapport à une valeur absolue se montre très féconde).
>
> – *Nicolas Bourbaki*
> *Topologie Générale, Book 3*

Challenge Problems: II

PROBLEM 2.1. Find $\mathrm{Aut}\,(\mathbb{C})$, the automorphism group of the field of complex numbers.

PROBLEM 2.2. Suppose that α is an irrational real number. Show that there exists an infinite number of rational numbers of the form p/q such that

$$|\alpha - p/q| < \frac{1}{\sqrt{5}q^2}.$$

PROBLEM 2.3. If c is a real number with $c > \sqrt{5}$, show that there exist irrational numbers α so that the inequality above fails when $\sqrt{5}$ is replaced by c.

PROBLEM 2.4. Let k be a positive integer. There is a longest finite sequence $x_1, x_2, \ldots, x_n$ from $\{1, 2, \ldots, k\}$ such that for no $i < j \leq n/2$ is $x_i, x_{i+1}, \ldots, x_{2i}$ a subsequence of $x_j, x_{j+1}, \ldots, x_{2j}$. Let $n(k)$ be the length of such a longest finite sequence. Show that $n(1) = 3$, $n(2) = 11$, and find $n(3)$.

PROBLEM 2.5. Suppose that a real number x has the decimal expansion $x = 0.101001000100001\ldots$. Show that x is irrational. Determine whether x is algebraic or transcendental.

2.1. Introduction

You have probably encountered the notion of distance in n-dimensional Euclidean space in previous mathematics courses. All that this involves is the repeated use of the Pythagorean theorem. If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are elements in $\mathbb{R}^n$, then we define

$$d(x, y) = ((x_1 - y_1)^2 + \cdots + (x_n - y_n)^2)^{1/2}.$$

Thus, if

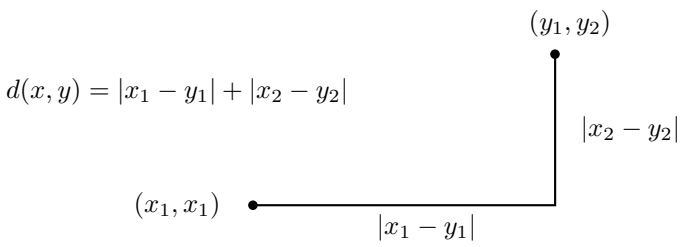
$$\|x\| = \left(\sum_{j=1}^{n} x_j^2 \right)^{1/2},$$

then $d(x, y) = \|x - y\|$.

In this chapter, we introduce the notion of a metric on a set. The word "metric" is simply a synonym for distance. As we shall see below, a metric on a set comes with a list of names attached to different subsets of the ambient space such as open set, closed set, etc. The open sets have some particular properties—so do the closed sets. These properties of subsets of metric spaces are special cases of similar properties for subsets of "topological spaces". The basic theory of topological spaces is developed in a project at the end of this chapter.

REMARK 2.1.1. Note that the metric d on $\mathbb{R}^n$ defined above is called the *usual metric on* $\mathbb{R}^n$.

It should be mentioned that this is not the only way to define the distance between two points in $\mathbb{R}^n$. For example, in $\mathbb{R}^2$, we could define $d_1(x, y) = |x_1 - y_1| + |x_2 - y_2|$. This is sometimes called the taxicab metric. This is what happens when you are driving on city streets and are not allowed to drive across peoples' lawns. The distance is illustrated below:



2.2. Definition and Basic Properties of Metric Spaces

DEFINITION 2.2.1. A *metric space* is a pair (X, d) where X is a set and $d : X \times X \longrightarrow \mathbb{R}$ is a map (the *metric*) satisfying the following properties.

(a) (Positive definite) For $x_1, x_2 \in X$, $d(x_1, x_2) \geq 0$ and $d(x_1, x_2) = 0$ if and only if $x_1 = x_2$.
(b) (Symmetric) For any $x_1, x_2 \in X$, we have $d(x_1, x_2) = d(x_2, x_1)$.

(c) (Triangle inequality) For any $x_1, x_2, x_3 \in X$, we have

$$d(x_1, x_2) \leq d(x_1, x_3) + d(x_3, x_2).$$

EXERCISE 2.2.2.

(i) Draw a triangle and figure out why the triangle inequality is so named.

(ii) Replace the triangle inequality by the inequality

$$d(x_1, x_2) \leq d(x_1, x_3) + d(x_2, x_3)$$

for any $x_1, x_2, x_3 \in X$. Show that symmetry follows from this version of the triangle inequality and positive definiteness.

EXERCISE 2.2.3. Show that the usual metric on $\mathbb{R}^n$ satisfies this definition.

EXERCISE 2.2.4. On $\mathbb{C}^n = \{z = (z_1, z_2, \ldots, z_n) \mid z_j \in \mathbb{C}\}$, we define

$$\|z\| = \left(\sum_{j=1}^{n} |z_j|^2 \right)^{1/2}$$

and, for $z, w \in \mathbb{C}^n$, we define $d(z, w) = \|z - w\|$. Show that d is a metric on $\mathbb{C}^n$.

EXERCISE 2.2.5. Let X be any nonempty set and, for $x_1, x_2 \in X$, define

$$d(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2, \\ 1 & \text{if } x_1 \neq x_2. \end{cases}$$

Show that d is a metric on X. This is called the *discrete metric*. It is designed to disabuse people of the notion that every metric looks like the usual metric on $\mathbb{R}^n$. The discrete metric is very handy for producing counterexamples.

EXAMPLE 2.2.6. Expanding on Remark 2.1.1, we introduce an important collection of metrics on $\mathbb{R}^n$.

Let p be a real number such that $p \geq 1$. For $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, we define

$$\|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}.$$

This is a key example for future developments. As usual, if $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, we define $d_p(x, y) = \|x - y\|_p$. To show that d_p is a metric on $\mathbb{R}^n$, we need the following inequality:

THEOREM 2.2.7 **(Hölder's inequality)**. *Suppose p, q are real numbers greater than 1 such that $1/p + 1/q = 1$. Suppose $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$. Then*

$$\sum_{k=1}^{n} |x_k y_k| \leq \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p} \left(\sum_{k=1}^{n} |y_k|^q \right)^{1/q}.$$

Proof. The proof is based on the following inequality. Suppose a and b are positive real numbers and p and q are as in the hypothesis of the theorem. Then $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$. This is proved with elementary calculus. Consider the function $y = x^{p-1}$, for $x \geq 0$. Then the inverse function is $x = y^{q-1}$, for $y \geq 0$. We have $\int_0^a x^{p-1}dx + \int_0^b y^{q-1}dy = \frac{a^p}{p} + \frac{b^q}{q}$. A look at the graphs of the functions reveals immediately that this sum is greater than or equal to ab where equality holds if and only if $a = b$.

Using this inequality, we get

$$\sum_{k=1}^{n} \frac{|x_k|}{||x||_p} \frac{|y_k|}{||y||_q} \leq \sum_{k=1}^{n} \frac{|x_k|^p}{p||x||_p^p} + \sum_{k=1}^{n} \frac{|y_k|^q}{q||y||_q^q} = 1/p + 1/q = 1.$$

EXERCISE 2.2.8. Now prove that d_p is a metric on $\mathbb{R}^n$. *Hint:* The triangle inequality is the only hard part. The proof of the triangle inequality depends on Hölder's inequality. To begin, observe that

$$||x + y||_p^p = \sum_i |x_i + y_i|^p \leq \sum_{i=1}^{n} |x_i + y_i|^{p-1}|x_i| + \sum_{i=1}^{n} |x_i + y_i|^{p-1}|y_i|.$$

Now apply Hölder's inequality.

EXERCISE 2.2.9. Note that Hölder's inequality works for $p, q > 1$. Prove the triangle inequality for the d_1 metric.

We also define a metric for $p = \infty$. That is, if $x = (x_1, x_2, \ldots, x_n)$, we set $\|x\|_\infty = \max_{1 \leq j \leq n} |x_j|$ and define $d_\infty(x, y) = \max_{1 \leq j \leq n} |x_j - y_j| = \|x - y\|_\infty$. It is easy to prove that d_∞ defines a metric on $\mathbb{R}^n$. The space $(\mathbb{R}^n, d_p)$ or alternatively $(\mathbb{R}^n, \|\cdot\|_p)$, $1 \leq p \leq \infty$, is denoted by $\ell_n^p(\mathbb{R})$. Note that, in our present notation, the norm symbol $\|\cdot\|$ on $\mathbb{R}^n$ should be relabeled $\|\cdot\|_2$.

EXERCISE 2.2.10. Show that everything we have just done for $\mathbb{R}^n$ can also be done for $\mathbb{C}^n$. This yields a collection of spaces $\ell_n^p(\mathbb{C})$.

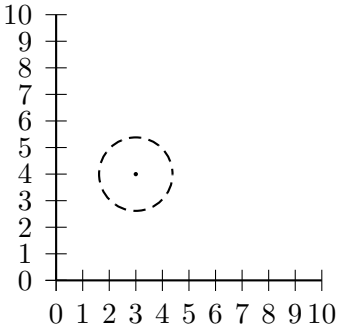
There is a lengthy menu attached to the study of metric spaces. For example, we need to deal with such concepts as open sets, closed sets, compact sets, accumulation points, isolated points, boundary points, interior, closure, and other things. To understand metric spaces fully, the reader must deal not only with these ideas, but with the relationships among them. Most of these ideas have a setting in the context of general topological spaces.

2.3. Topology of Metric Spaces

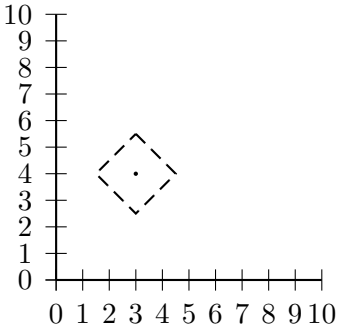
A fundamental notion in the study of metric spaces is that of an open ball.

DEFINITION 2.3.1. Suppose that (X, d) is a metric space and $x_0 \in X$. If $r \in \mathbb{R}$, with $r > 0$, the *open ball of radius r around x_0* is the subset of X defined by $B_r(x_0) = \{x \in X \mid d(x, x_0) < r\}$. Similarly, we define the *closed ball of radius r around x_0* by $\overline{B}_r(x_0) = \{x \in X \mid d(x, x_0) \leq r\}$.

EXAMPLE 2.3.2. In $\mathbb{R}^2$, with the usual metric, a ball of radius $3/2$ around the point $(3, 4)$ looks like this:

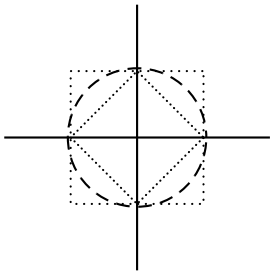


EXAMPLE 2.3.3. In $\mathbb{R}^2$, a ball of radius $3/2$ around the point $(3, 4)$ in the d_1 metric looks like this:



DEFINITION 2.3.4. Suppose that V is a vector space (see Appendix B) with a metric d. The *unit ball in* V is the ball of radius 1 with center at $\mathbf{0}$, that is, $B_1(\mathbf{0})$.

We want to analyze the nature of the *unit ball* in $\ell_n^p(\mathbb{R})$, that is, the set of all points $x \in \mathbb{R}^n$ such that $\|x\|_p < 1$. For the moment, let us take $n = 2$ and consider the cases of $\ell_2^1(\mathbb{R})$, $\ell_2^2(\mathbb{R})$, and $\ell_2^\infty(\mathbb{R})$. The pictures of the unit balls in these spaces are shown below. This leads to an interesting diagram:



EXERCISE 2.3.5. If $1 \leq p < q$, show that the unit ball in $\ell_n^p(\mathbb{R})$ is contained in the unit ball in $\ell_n^q(\mathbb{R})$.

EXERCISE 2.3.6. Consider the set of all points in $\mathbb{R}^2$ which lie outside the unit ball in $\ell_2^1(\mathbb{R})$ and inside the unit ball in $\ell_2^\infty(\mathbb{R})$. Does every point in this region lie on the perimeter of the unit ball in $\ell_2^p(\mathbb{R})$ for some p between 1 and ∞? Do the same problem for $\ell_n^p(\mathbb{R})$.

Next, we look at open sets.

DEFINITION 2.3.7. Let (X, d) be a metric space and suppose that $A \subseteq X$. The set A is an *open set* in X if, for each $a \in A$, there is an $r > 0$ such that $B_r(a) \subseteq A$.

Notice that the radius r depends on the point a. Also, observe that the empty set $\varnothing$ and the whole space X are both open sets.

EXERCISE 2.3.8. Prove that, for any $x_0 \in X$ and any $r > 0$, the "open ball" $B_r(x_0)$ is open. So now we can legitimately call an "open" ball an open set.

EXERCISE 2.3.9. Prove that the following are open sets:

(i) the "first quadrant", $\{(x, y) \in \mathbb{R}^2 \mid x > 0 \text{ and } y > 0\}$, in the usual metric;
(ii) any subset of a discrete metric space.

Open sets behave nicely under certain set-theoretic operations.

THEOREM 2.3.10.

(i) *If $\{A_j\}_{j \in J}$ is a family of open sets in a metric space (X, d), then*

$$\bigcup_{j \in J} A_j$$

is an open set in X.

(ii) *If $A_1, A_2, \ldots, A_n$ are open sets in a metric space (X, d), then*

$$\bigcap_{j=1}^{n} A_j$$

is an open set in X.

Proof. (i) Suppose that $x \in \bigcup_{j \in J} A_j$. Then $x \in A_k$ for some $k \in J$. Since A_k is open, there is a real number $r > 0$ such that $B_r(x) \subseteq A_k$. But then, $B_r(x) \subseteq \bigcup_{j \in J} A_j$.

(ii) Suppose $x \in \bigcap_{j=1}^n A_j$. Then $x \in A_j$ for each $j = 1, 2, \ldots, n$. Since A_j is open, for each j, there exists a radius r_j such that $B_{r_j}(x) \subseteq A_j$. Let $r = \min_{1 \le j \le n}\{r_j\}$. Then $r > 0$ and $B_r(x) \subseteq \bigcap_{j=1}^n A_j$.

We can now say that open sets are closed under arbitrary union and finite intersection.

DEFINITION 2.3.11. Let (X, d) be a metric space and let $\mathcal{O}$ be the collection of all open sets in X relative to d. The pair $(X, \mathcal{O})$ is an example of a topological space (see the project in Section 2.7.1) and the collection $\mathcal{O}$ is called a *topology* on X.

EXERCISE 2.3.12.

(i) There can be problems with infinite intersections. For example, let $A_n = B_{1/n}((0,0))$ in $\mathbb{R}^2$ with the usual metric. Show that

$$\bigcap_{n=1}^{\infty} A_n$$

is not open.

(ii) Find an infinite collection of distinct open sets in $\mathbb{R}^2$ with the usual metric whose intersection is a nonempty open set.

Thus infinite intersections of open sets may or may not be open.
If there are open sets in a metric space, can closed sets be far behind?

DEFINITION 2.3.13. Let (X, d) be a metric space and suppose that $A \subseteq X$. We say that A is a *closed set* in X if ${}^cA$ is open in X. (Recall that ${}^cA = X \setminus A$ is the complement of A in X.)

EXERCISE 2.3.14. Show that the following are closed sets:

(i) the x-axis in $\mathbb{R}^2$ with the usual metric;
(ii) the whole space X in any metric space;
(iii) the empty set in any metric space;
(iv) a single point in any metric space;
(v) any subset of a discrete metric space.

EXERCISE 2.3.15. Show that a "closed ball" $\overline{B}(x_0)$ is a closed set.

EXERCISE 2.3.16. Show that $\mathbb{Q}$ as a subset of $\mathbb{R}$ with the usual metric is neither open nor closed in $\mathbb{R}$. Of course, if the metric space is simply $\mathbb{Q}$ with the usual metric, then $\mathbb{Q}$ is both open and closed in $\mathbb{Q}$.

Here is a basic theorem about closed sets.

THEOREM 2.3.17.

(i) *Suppose that (X, d) is a metric space and that $\{A_j\}_{j \in J}$ is a collection of closed sets in X. Then*

$$\bigcap_{j \in J} A_j$$

is a closed set in X.

(ii) *If $A_1, A_2, \ldots, A_n$ are closed sets in X, then*

$$\bigcup_{j=1}^{n} A_j$$

is a closed set in X.

Proof. Use Theorem 2.3.10 and DeMorgan's laws.

So, a set is closed iff its complement is open, and a set is open iff its complement is closed. However, most of the time, sets in a metric space are neither open nor closed. There is a different way to characterize closed sets. First, we need the notion of an accumulation point. From here on, we shall simply refer to a metric space X and suppress the notation d for the metric.

DEFINITION 2.3.18. Suppose that A is a subset of a metric space X. A point $x_0 \in X$ is an *accumulation point* of A if, for every $r > 0$, we have $(B_r(x_0) \setminus \{x_0\}) \cap A \neq \varnothing$.

Thus, if x_0 is an accumulation point of A, there are points of A (other than x_0) that are arbitrarily close to x_0. Note that x_0 may or may not be an element of A. For example, for $\mathbb{R}$ with the usual metric, 1 and 0 are accumulation points of the open interval $(0, 1)$ as well as all of the points in the interval itself.

DEFINITION 2.3.19. Suppose that A is a subset of a metric space X. A point $x_0 \in A$ is an *isolated point* of A if there is an $r > 0$ such that $B_r(x_0) \cap A = \{x_0\}$.

DEFINITION 2.3.20. Suppose that A is a subset of a metric space X. A point $x_0 \in X$ is a *boundary point* of A if, for every $r > 0$, $B_r(x_0) \cap A \neq \varnothing$ and $B_r(x_0) \cap {}^cA \neq \varnothing$. The *boundary of* A is the set of boundary points of A and is denoted by ∂A.

We need some examples.

EXAMPLES 2.3.21.

(i) Let $A = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 < 1\}$. We take the usual metric on $\mathbb{R}^3$. The set of accumulation points of A is $B^3 = \{(x, y, z) \mid x^2 + y^2 + z^2 \leq 1\}$ and is called *the closed unit ball* in $\mathbb{R}^3$ with respect to the usual metric. The set A has no isolated points, and $\partial A = S^2 = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}$. The set S^2 is called the *2-sphere* in $\mathbb{R}^3$ with respect to the usual metric.

(ii) Let $A = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + x_2^2 + \cdots + x_n^2 < 1\}$. We take the usual metric in $\mathbb{R}^n$. The set of accumulation points of A is $B^n = \{(x_1, x_2, \ldots, x_n) \mid x_1^2 + x_2^2 + \cdots + x_n^2 \leq 1\}$. The set A is called *the open unit ball* with respect to the usual metric and the set B^n is called *the closed unit ball* in $\mathbb{R}^n$ with respect to the usual metric. The set A has no isolated points and $\partial A = S^{n-1} = \{(x_1, x_2, \ldots, x_n) \mid x_1^2 + x_2^2 + \cdots + x_n^2 = 1\}$. The set S^{n-1} is called the *n-sphere* in $\mathbb{R}^n$ with respect to the usual metric.

(iii) Let $A = \mathbb{Q} \subseteq \mathbb{R}$ with the usual metric. Then every point in $\mathbb{R}$ is an accumulation point of A, the set A has no isolated points, and $\partial A = \mathbb{R}$.

(iv) If A is any subset of a discrete metric space X, then A has no accumulation points. Every point in A is an isolated point, and $\partial A = \varnothing$.

Accumulation points furnish another means of identifying closed sets.

THEOREM 2.3.22. *Suppose A is a subset of a metric space X. Then A is closed iff A contains all its accumulation points.*

Proof. If A is the empty set, then A has no accumulation points. Suppose that A is a nonempty closed set and that x_0 is an accumulation point of A. If $x_0 \notin A$, then $x_0 \in {}^cA$, which is open. Hence, there is an $r > 0$ such that $B_r(x_0) \subseteq {}^cA$, and this contradicts the definition of accumulation point. Conversely, suppose that A contains all its accumulation points and that $x_0 \in {}^cA$. Then x_0 is not an accumulation point of A, and hence there exists $r > 0$ such that $B_r(x_0) \cap A = \varnothing$. This means that ${}^cA$ is open, and so A is closed.

In a discrete metric space any subset is both open and closed. This is not generally the case. For example, in the case of $\ell_n^p(\mathbb{R})$ and $\ell_n^p(\mathbb{C})$, most subsets are neither open nor closed.

EXERCISE 2.3.23. Find an uncountable number of subsets of $\ell_n^p(\mathbb{R})$ and $\ell_n^p(\mathbb{C})$ which are neither open nor closed.

If a set A in a metric space X is not closed, we can do something about it.

DEFINITION 2.3.24. Suppose that A is a subset of a metric space X. The *closure of* A is the intersection of all the closed sets which contain A.

The closure of any set A exists, since there are always closed sets which contain A, for example X. The closure of A is a closed set since it is the intersection of closed sets. So the closure of A is the "smallest" closed set that contains A. We denote the closure of a set A by $\overline{A}$. Obviously, $A \subseteq \overline{A}$ and $A = \overline{A}$ iff A is closed.

EXAMPLES 2.3.25.

(i) Let $A = \{(x, y, z) \in \mathbb{R}^3 \mid x > 0, y > 0, z > 0\}$. If $\mathbb{R}^3$ has the usual metric, then $\overline{A} = \{(x, y, z) \in \mathbb{R}^3 \mid x \geq 0, y \geq 0, z \geq 0\}$.

(ii) Let $\mathbb{Q}^n = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid x_j \in \mathbb{Q} \text{ for } 1 \leq j \leq n\}$. If $\mathbb{R}^n$ has the usual metric, then $\overline{\mathbb{Q}^n} = \mathbb{R}^n$.

(iii) Let X be a discrete metric space and let A be any subset of X. Then $\overline{A} = A$.

It should not come as a surprise that the notions of closure and accumulation point are intimately related.

EXERCISE 2.3.26. Suppose that A is a subset of a metric space X. Show that $\overline{A} = A \cup \{\text{accumulation points of } A\}$.

EXERCISE 2.3.27. Suppose A is a subset of a metric space X. Prove or disprove that $\overline{A} = A \cup \partial A$.

EXERCISE 2.3.28. Let X be a metric space and let $x_0 \in X$. Suppose that $r > 0$. Prove or disprove that $\overline{B_r(x_0)} = \overline{B}_r(x_0)$.

REMARK 2.3.29. Suppose that (X, d) is a metric space and Y is a nonempty subset of X. Then (Y, d) is a metric space.

EXERCISE 2.3.30. Show that the topology on Y is the *relative topology* from X. That is, show that a subset of Y is open with respect to d iff it is the intersection of an open subset of X with Y.

EXERCISE 2.3.31. The definitions and notations for this exercise may be found in Appendix B.

(i) Consider the set of 2×2 matrices over $\mathbb{R}$, that is, $M_2(\mathbb{R})$. Make this into a metric space by identifying it with $\mathbb{R}^4$ with the usual metric. Show that $GL_2(\mathbb{R})$ is an open subset of $M_2(\mathbb{R})$ and that $\overline{GL_2(\mathbb{R})} = M_2(\mathbb{R})$.
(ii) Show that $SL_2(\mathbb{R})$ is a closed subset of $GL_2(\mathbb{R})$.

EXERCISE 2.3.32. Let A be a subset of a metric space X and let x_0 be an isolated point of A. Show that x_0 is in the boundary of A if and only if x_0 is an accumulation point ${}^cA$.

EXERCISE 2.3.33. As usual, let $\mathbb{R}$ be an ordered field with the least upper bound property. Give $\mathbb{R}$ the discrete metric. Show that $\mathbb{R}$ is still an ordered field with the least upper bound property but that neither the rational nor irrational numbers are dense in $\mathbb{R}$. Determine what other relevant properties of $\mathbb{R}$ with the usual metric do not hold with the discrete metric.

Corresponding to the notion of closure is the idea of the *interior* of a set.

DEFINITION 2.3.34. Let A be a subset of a metric space X. The *interior* of A is the union of all open sets which are contained in A.

The interior of A is the largest open set contained in A. We denote the interior of A by A°. Obviously $A^\circ \subseteq A$ and $A^\circ = A$ iff A is open.

EXAMPLES 2.3.35.

(i) Let $X = \mathbb{R}^3$ with the usual metric and let $A = \{(x, y, z) \mid z \geq 0\}$. Then $A^\circ = \{(x, y, z) \mid z > 0\}$.
(ii) Let X be a discrete metric space and let A be any subset of X. Then $A^\circ = A$ and $\overline{A} = A$, so that $A = A^\circ = \overline{A}$.

EXERCISE 2.3.36. Show that, in the usual metric on $\mathbb{R}$, the interior of $\mathbb{Q}$ is empty, that is, $\mathbb{Q}^\circ = \varnothing$, but the interior of $\bar{\mathbb{Q}}$ is $\mathbb{R}$, that is, $(\bar{\mathbb{Q}})^\circ = \mathbb{R}$.

EXERCISE 2.3.37. Look at combinations of interior, closure, and boundary and determine how many different possibilities result. For this exercise only, let "I" stand for interior, let "B" stand for boundary, and let "C" stand for closure. Let X be a metric space and let $A \subseteq X$. How many possible sets can be made from A with these operations? For example,

$I(I(A)) = I(A)$ but $C(I(A))$ is not necessarily A. Is it $C(A)$? Explore all possibilities of applying combinations of I, C, and B. Hint: There are only a finite number.

Another important concept in the theory of metric spaces is that of diameter.

DEFINITION 2.3.38. Let A be a nonempty subset of a metric space X. The *diameter of* A is

$$\operatorname{diam}(A) = \sup_{x,y \in A} d(x, y).$$

EXERCISE 2.3.39.

(i) Show that the diameter of a set is 0 iff the set consists of a single point.
(ii) Suppose A is a nonempty subset of a metric space X. Show that $\operatorname{diam}(A) = \operatorname{diam}(\overline{A})$.

DEFINITION 2.3.40. Let A be a nonempty subset of $\mathbb{R}^n$. We say that A is *convex* if, given any two points $P, Q \in A$, the set of points $\{(1-t)P + tQ \mid t \in \mathbb{R},\ 0 \leq t \leq 1\}$ is a subset of A.

EXAMPLE 2.3.41. The unit ball B^n contained in $\mathbb{R}^n$, in the usual metric, is a convex set.

EXERCISE 2.3.42. Show that the unit ball in $\ell_n^p(\mathbb{R})$, for $1 \leq p \leq \infty$, is a convex set in $\mathbb{R}^n$.

DEFINITION 2.3.43. Let A be a subset of $\mathbb{R}^n$ with the usual metric. The *convex hull* of A is the intersection of all convex sets containing A. The *closed convex hull* of A is the intersection of all closed convex sets containing A.

EXERCISE 2.3.44. Let A be a nonempty subset of $\mathbb{R}^n$ with the usual metric and let C be the convex hull of A.

(i) Prove or disprove the following statement. The closed convex hull of A is $\overline{C}$.
(ii) Show that the diameter of A is the diameter of C.

REMARK 2.3.45. The concept of convex set in $\mathbb{R}^n$ does not involve a metric in $\mathbb{R}^n$. However, a particular metric is often used to define subsets of $\mathbb{R}^n$ which may or may not be convex.

EXERCISE 2.3.46.

(i) Describe the closed convex hull of the unit ball in $\ell_n^p(\mathbb{R})$ for $1 \leq p \leq \infty$.
(ii) Suppose $0 < p < 1$. For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, define

$$\|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{\frac{1}{p}}.$$

Define $S_p = \{x \in \mathbb{R}^n \mid \|x\|_p \leq 1\}$. Determine whether S_p is convex. If not, find the closed convex hull of S_p.

EXAMPLE 2.3.47. We now branch out in a slightly different direction. Suppose that X is a set and that $F = \mathbb{R}$ or $\mathbb{C}$. Denote by $\mathcal{B}(X, F)$ the set of all functions from X to F which are bounded. Thus, $f \in \mathcal{B}(X, F)$ iff there is a real number M such that $|f(x)| \leq M$ for all $x \in X$. For $f, g \in \mathcal{B}(X, F)$, we define

$$d(f, g) = \sup_{x \in X} |f(x) - g(x)|.$$

It follows easily from the definition of sup that d is a metric on $\mathcal{B}(X, F)$. In this example, an open ball of radius r around a function f is the collection of all functions which lie within an "r-strip" around f.

EXERCISE 2.3.48.

(i) Let $F = \mathbb{R}$ or $\mathbb{C}$. Show that $\mathcal{B}(X, F)$, with d as defined above, is a metric space.

(ii) For $f, g \in \mathcal{B}(X, F)$, define $(f + g)(x) = f(x) + g(x)$ and $(fg)(x) = f(x)g(x)$. Also, for $\alpha \in F$ define $(\alpha f)(x) = \alpha f(x)$. Show that, with these operations, $\mathcal{B}(X, F)$ is a commutative algebra with 1 over F. Of course, scalar multiplication is simply multiplication by a constant function.

This is a step up in our examples of metric spaces. While previous examples are important, spaces of functions are the most significant examples of metric spaces in analysis.

2.4. Limits and Continuous Functions

The consideration of function spaces leads us to the notion of continuous function. Before we define continuous functions, we introduce the idea of limit.

DEFINITION 2.4.1. Suppose $(a_n)_{n \in \mathbb{N}}$ is a sequence of points in a metric space X. We say that a point $L \in X$ is the *limit* of the sequence $(a_n)_{n \in \mathbb{N}}$ as n goes to infinity if, for any $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that $d(a_n, L) < \varepsilon$ whenever $n \geq N$. When the limit exists, we say that $(a_n)_{n \in \mathbb{N}}$ *converges to* L and write

$$\lim_{n \to \infty} a_n = L.$$

Sometimes, we simply say that $(a_n)_{n \in \mathbb{N}}$ *converges in* X without mentioning L explicitly.

As in Chapter 1, we have a concept of Cauchy sequences in a metric space.

DEFINITION 2.4.2 (See Definitions 1.5.2 and 1.6.12). Let X be a metric space and let $(a_n)_{n \in \mathbb{N}}$ be a sequence in X. We say that $(a_n)_{n \in \mathbb{N}}$ is a *Cauchy sequence* if, for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(a_n, a_m) < \varepsilon$ whenever $n, m \geq N$.

It may be that a sequence in a metric space is a Cauchy sequence even though it does not converge. For example, as we observed in Chapter 1, Cauchy sequences in $\mathbb{Q}$ with the usual metric do not necessarily converge in $\mathbb{Q}$. This leads us to the following exercise.

EXERCISE 2.4.3. Suppose that X is a metric space and that the sequence $(a_n)_{n\in\mathbb{N}}$ converges in X. Show that, for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(a_n, a_m) < \varepsilon$ whenever $n, m \geq N$. Thus, a convergent sequence is a Cauchy sequence.

EXERCISE 2.4.4. Let $(a_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in a discrete metric space X. Show that there exists $N \in \mathbb{N}$ such that $d(a_n, a_m) = 0$, that is, $a_n = a_m$, for all $n, m \geq N$. Such a sequence is called *eventually constant*. Note that an eventually constant sequence in any metric space is convergent and, in fact, it converges to the eventual constant.

There is a standard litany associated to the notions of convergent sequence and Cauchy sequence. For example, from Exercise 2.4.3, we see that in any metric space, a convergent sequence is a Cauchy sequence. In $\mathbb{R}$ or $\mathbb{C}$ with the usual metric, every Cauchy sequence converges. In $\mathbb{Q}$ with the usual metric, many Cauchy sequences do not converge. The best kinds of metric spaces (at least our favorite kinds) are the ones in which "convergent sequence" and "Cauchy sequence" are synonymous.

DEFINITION 2.4.5. Suppose that X is a metric space. We say that X is a *complete metric space* if every Cauchy sequence in X converges.

EXAMPLES 2.4.6. The following metric spaces are complete. (If this seems repetitive, don't worry about it.) The proofs of (i) and (ii) are in Chapter 1:

(i) $\mathbb{R}$ with the usual metric;
(ii) $\mathbb{C}$ with the usual metric;
(iii) any discrete metric space.

The rational numbers $\mathbb{Q} \subseteq \mathbb{R}$ are not complete in the usual metric, but they are complete in the discrete metric.

EXERCISE 2.4.7. Prove that a closed subset of a complete metric space is a complete metric space with the inherited metric.

EXERCISE 2.4.8. Show that, for $1 \leq p \leq \infty$, the spaces $\ell_n^p(\mathbb{R})$ and $\ell_n^p(\mathbb{C})$ are complete metric spaces.

We now turn to an investigation of convergence and completeness in the spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ with the usual metrics. We proceed in a fashion very similar to that of $\mathbb{R}$. One big difference is that, since there is no notion of order in $\mathbb{R}^n$, the idea of monotonicity has no meaning. However, we will use it one variable at a time. The following lemmas and theorems for $\mathbb{R}^n$ and $\mathbb{C}^n$ will be proved for $\mathbb{R}^n$ and left as exercises for $\mathbb{C}^n$.

LEMMA 2.4.9. *Every bounded sequence in $\mathbb{R}^n$ or $\mathbb{C}^n$ with the usual metric has a convergent subsequence.*

Proof. Let $(a_m)_{m\in\mathbb{N}}$ be a bounded sequence in $\mathbb{R}^n$. Write $a_m = (a_{m,1}, a_{m,2}, \ldots, a_{m,n})$. We prove the lemma by induction on n. For $n = 1$, this is the content of Lemma 1.6.10. Assume the lemma is true for $n-1$. Let a'_m be the $(n-1)$-tuple $(a_{m,1}, a_{m,2}, \ldots, a_{m,n-1})$. Then $(a'_m)_{m\in\mathbb{N}}$ is a bounded sequence in $\mathbb{R}^{n-1}$. By the induction hypothesis, $(a'_m)_{m\in\mathbb{N}}$ has a convergent subsequence in $\mathbb{R}^{n-1}$. Label this convergent subsequence $(a'_{m_j})_{j\in\mathbb{N}}$ and then take the corresponding subsequence in the sequence $(a_{m_j,n})_{j\in\mathbb{N}}$. Now the sequence $(a_{m_j,n})_{j\in\mathbb{N}}$ is a bounded sequence in $\mathbb{R}$ and hence has a convergent subsequence which we shall not name. Again, taking the subsequence and choosing the corresponding subsequence of $(a_{m_j})_{j\in\mathbb{N}}$, we get a convergent subsequence of the original sequence $(a_m)_{m\in\mathbb{N}}$.

EXERCISE 2.4.10.

(i) For practice, carry out the above proof in $\mathbb{C}^n$.
(ii) Prove the above lemma by proceeding coordinate by coordinate. You will notice that the indexing gets quite messy.

THEOREM 2.4.11 **(Bolzano-Weierstrass)**. *If A is a bounded infinite subset of $\mathbb{R}^n$ or $\mathbb{C}^n$, then A has an accumulation point.*

Proof. Since A is infinite, there exists a sequence $(x_k)_{k\in\mathbb{N}}$, where $x_k \neq x_j$ if $k \neq j$. Then $(x_k)_{k\in N}$ is a bounded sequence in $\mathbb{R}^n$ and by Lemma 2.4.9 has a convergent subsequence. If this subsequence converges to x_0, then x_0 is an accumulation point of A.

REMARK 2.4.12. As in Example 2.3.47, we let $\mathcal{B}(X, F)$ denote either $\mathcal{B}(X, \mathbb{R})$ or $\mathcal{B}(X, \mathbb{C})$. There are two types of convergence to be discussed in this space. The first is called *uniform convergence*, that is, convergence with respect to the metric defined in the example. In this case, a sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{B}(X, F)$ is convergent if there exists $f \in \mathcal{B}(X, F)$ such that given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\sup_{x\in X} |f_n(x) - f(x)| < \varepsilon$$

for $n \geq N$. Then, for a fixed $x_0 \in X$, the sequence $(f_n(x_0))_{n\in\mathbb{N}}$ converges to $f(x_0)$. Thus uniform convergence of a sequence $(f_n)_{n\in\mathbb{N}}$ implies pointwise convergences as well.

Observe that even if a sequence does not converge uniformly, it may still have a pointwise limit in $\mathcal{B}(X, F)$. For example, define the following sequence of functions in $\mathcal{B}([0, 1], \mathbb{R})$:

$$f_n(x) = \begin{cases} 2n^2 x & \text{if } 0 \leq x \leq \frac{1}{2n}, \\ -2n^2(x - \frac{1}{n}) & \text{if } \frac{1}{2n} \leq x \leq \frac{1}{n}, \\ 0 & \text{if } \frac{1}{n} \leq x \leq 1. \end{cases}$$

Then, for each $x \in [0,1]$, we have $\lim_{n\to\infty} f_n(x) = 0$. So, the sequence $(f_n)_{n\in\mathbb{N}}$ converges pointwise to the function $f(x) = 0$, for every $x \in [0,1]$. However, it is clear that this convergence is not uniform. Note that all the functions f_n as well as the limit function f are continuous by elementary calculus.

EXERCISE 2.4.13. Show that a Cauchy sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{B}(X,F)$ converges pointwise to some function f.

We now ask whether a Cauchy sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{B}(X,F)$ converges uniformly to its pointwise limit f.

THEOREM 2.4.14. *The spaces $\mathcal{B}(X,\mathbb{R})$ and $\mathcal{B}(X,\mathbb{C})$ are complete metric spaces.*

Proof. As above, we let $\mathcal{B}(X,F) = \mathcal{B}(X,\mathbb{R})$ or $\mathcal{B}(X,\mathbb{C})$. Suppose that $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{B}(X,F)$ and denote by f the pointwise limit of $(f_n)_{n\in\mathbb{N}}$. We want to show that the sequence $(f_n)_{n\in\mathbb{N}}$ converges uniformly to f. To this end, choose $\varepsilon > 0$ and $N \in \mathbb{N}$ such that $\sup_{x\in X} |f_n(x) - f_m(x)| < \varepsilon/2$ when $n, m \geq N$. Fix $x_0 \in X$ and choose an integer $N(x_0) \geq N$ such that $|f_{N(x_0)}(x_0) - f(x_0)| < \varepsilon/2$. Then

$$|f_n(x_0) - f(x_0)| \leq |f_n(x_0) - f_{N(x_0)}(x_0)| + |f_{N(x_0)}(x_0) - f(x_0)| < \varepsilon$$

if $n \geq N$.

To complete the proof, we must show that the function f is bounded, that is, $f \in \mathcal{B}(X,F)$. But, from the above inequality, it follows that $|f(x)| < |f_n(x)| + \varepsilon$ for all $x \in X$.

We now turn to one of the most important ideas about functions on metric spaces.

DEFINITION 2.4.15. Let (X,d) and (X',d') be metric spaces. A function $f : X \longrightarrow X'$ is *continuous at the point* $x_0 \in X$ if, for any $\varepsilon > 0$, there is a $\delta > 0$ such that $d'(f(x), f(x_0)) < \varepsilon$ whenever $x \in X$ and $d(x,x_0) < \delta$.

This is the old familiar ε-δ definition. It is simply the statement that

$$\lim_{x\to x_0} f(x) = f(x_0).$$

More generally, we say that

$$\lim_{x\to x_0} f(x) = L$$

for some $L \in X'$ if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $d'(f(x), L) < \varepsilon$ whenever $0 < d(x,x_0) < \delta$.

EXERCISE 2.4.16. Suppose that X and X' are metric spaces as above and that $x_0 \in X$. Show that f is continuous at x_0 iff for every sequence $(x_n)_{n\in\mathbb{N}}$ in X which converges to x_0 in X, we have

$$\lim_{n\to\infty} f(x_n) = f(x_0)$$

in X'.

Note that another way of saying that f is continuous at x_0 is the following. Given $\varepsilon > 0$, there exists $\delta > 0$ such that $f(B_\delta(x_0)) \subseteq B_\varepsilon(f(x_0))$.

In discussing continuity, one must be careful about the domain of the function. For example, define $f : \mathbb{R} \longrightarrow \mathbb{R}$ by the equation

$$f(x) = \begin{cases} 0 & \text{if } x \notin \mathbb{Q}, \\ 1 & \text{if } x \in \mathbb{Q}. \end{cases}$$

Then, f is not continuous at any point of $\mathbb{R}$. However, suppose we restrict f to be a function from $\mathbb{Q}$ to $\mathbb{Q}$. This means that $f(x) = 1$ on $\mathbb{Q}$ and is continuous at every point of $\mathbb{Q}$.

EXERCISE 2.4.17. Define $f : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} 1/q & \text{if } x = p/q \text{ (reduced to lowest terms, } x \neq 0), \\ 0 & \text{if } x = 0 \text{ or } x \notin \mathbb{Q}. \end{cases}$$

Show that f is continuous at 0 and any irrational point. Show that f is not continuous at any nonzero rational point.

Continuity is called a *pointwise property* or *local property* of a function f; that is, as in Exercise 2.4.17, a function may be continuous at some points, but not at others. We often deal with functions $f : X \longrightarrow X'$ which are continuous at every point of X. In this case, we simply say that f is *continuous* without reference to any particular point.

THEOREM 2.4.18. *Suppose that (X, d) and (X', d') are metric spaces. Then a function $f : X \longrightarrow X'$ is continuous iff for any open set $V \subseteq X'$, the set $f^{-1}(V)$ is an open set in X.*

Proof. First suppose that f is continuous. Let V be an open set in X'. Suppose $x_0 \in f^{-1}(V)$. Take $\varepsilon > 0$ such that $B_\varepsilon(f(x_0)) \subseteq V$. Then there exists $\delta > 0$ such that $f(B_\delta(x_0)) \subseteq B_\varepsilon(f(x_0))$, and so $B_\delta(x_0) \subseteq f^{-1}(B_\varepsilon(f(x_0))) \subseteq f^{-1}(V)$. So $f^{-1}(V)$ is open.

The second half of the proof is easy. You do it.

EXERCISE 2.4.19.

(i) Let X and X' be metric spaces and assume that X has the discrete metric. Show that any function $f : X \longrightarrow X'$ is continuous.

(ii) Let $X = \mathbb{R}$ with the usual metric and let $f : X \longrightarrow X$ be a polynomial function. Show that f is continuous.

(iii) Let $X = \mathbb{R}$ with the usual metric and let $X' = \mathbb{R}$ with the discrete metric. Describe all continuous functions from $X \longrightarrow X'$.

REMARK 2.4.20. In a general topological space (see the project in Section 2.7.1), the definition of continuity is ordinarily stated by saying that the inverse image of an open set is open (see Definition 2.7.12). However, in metric spaces, the definition of continuity using limits is often more useful.

DEFINITION 2.4.21. A subset A of a metric space X is *bounded* if there exists a point $x \in X$ and $r > 0$ such that $A \subseteq B_r(x)$.

EXERCISE 2.4.22. Suppose that (X, d) and (X', d') are metric spaces and that $f : X \longrightarrow X'$ is continuous. For each of the following statements, determine whether or not it is true. If the assertion is true, prove it. If it is not true, give a counterexample.

(i) If A is an open subset of X, then $f(A)$ is an open subset of X'.
(ii) If B is a closed subset of X', then $f^{-1}(B)$ is a closed subset of X.
(iii) If A is a closed subset of X, then $f(A)$ is a closed subset of X'.
(iv) If A is a bounded subset of X, then $f(A)$ is a bounded subset of X'.
(v) If B is a bounded subset of X', then $f^{-1}(B)$ is a bounded subset of X.
(vi) If $A \subseteq X$ and x_0 is an isolated point of A, then $f(x_0)$ is an isolated point of $f(A)$.
(vii) If $A \subseteq X$, $x_0 \in A$, and $f(x_0)$ is an isolated point of $f(A)$, then x_0 is an isolated point of A.
(viii) If $A \subseteq X$ and x_0 is an accumulation point of A, then $f(x_0)$ is an accumulation point of $f(A)$.
(ix) If $A \subseteq X$, $x_0 \in X$, and $f(x_0)$ is an accumulation point of $f(A)$, then x_0 is an accumulation point of A.
(x) Do any of your answers to the above questions change if we assume X and/or X' is complete?

DEFINITION 2.4.23. Let (X, d) and (X', d') be metric spaces. A continuous function $f : X \longrightarrow X'$ is a *homeomorphism* if

(a) f is a bijection and
(b) the function f^{-1} is also continuous.

THEOREM 2.4.24. *Suppose $1 \leq p < q \leq \infty$. Then the identity map $I(x) = x$ from $\ell_n^p(\mathbb{R})$ to $\ell_n^q(\mathbb{R})$ is a homeomorphism.*

Proof. From our previous discussion, it is clear that if $p < q$, then the unit ball in $\ell_n^p(\mathbb{R})$ is contained in the unit ball in $\ell_n^q(\mathbb{R})$. Take $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and suppose $\max_{1 \leq i \leq n}(|x_i|) \leq 1$. Then

$$\left|\frac{x_1}{n}\right| + \cdots + \left|\frac{x_n}{n}\right| \leq 1.$$

This shows that the ball of radius $1/n$ in the ℓ_n^∞ metric is contained in the ball of radius 1 in the ℓ_n^1 metric. In particular, this last fact shows that if we take the unit ball in $\ell_n^q(\mathbb{R})$ and multiply each coordinate by a factor of $1/n$, then the resulting set of points is contained in the unit ball in $\ell_n^p(\mathbb{R})$. This is enough to show that I is a homeomorphism.

EXERCISE 2.4.25. Show that $\ell_n^p(\mathbb{C})$ and $\ell_n^q(\mathbb{C})$ are homeomorphic.

DEFINITION 2.4.26. A homeomorphism $f : X \longrightarrow X'$ is an *isometry* if

$$d'(f(x_1), f(x_2)) = d(x_1, x_2)$$

for all $x_1, x_2 \in X$. The metric spaces X and X' are said to be *isometric* if there exists an isometry between them.

EXERCISE 2.4.27. Suppose that, instead, we had defined an isometry to be a bijection $f : X \longrightarrow X'$ such that $d'(f(x_1), f(x_2)) = d(x_1, x_2)$ for all $x_1, x_2 \in X$. Show that with this definition any isometry is a homeomorphism.

EXERCISE 2.4.28. Let $X = \mathbb{R}$ with the discrete metric and let $X' = \mathbb{R}$ with the usual metric. Define $f : X \longrightarrow X'$ by $f(x) = x$. Show that f is a continuous bijection which is not a homeomorphism.

EXERCISE 2.4.29. Let (X, d) be a metric space. Let G be the collection of all homeomorphisms from X to X. Prove that, under composition of functions, G is a group and the collection of all isometries is a subgroup of G.

EXERCISE 2.4.30.

(i) Suppose $f : \mathbb{R} \longrightarrow \mathbb{R}$ satisfies $|f(x_1) - f(x_2)| = |x_1 - x_2|$ for all $x_1, x_2 \in \mathbb{R}$. Show that f is a surjection and determine its explicit form.
(ii) Suppose $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ satisfies $\|f(x_1) - f(x_2)\| = \|x_1 - x_2\|$ for all $x_1, x_2 \in \mathbb{R}^n$, where $\|\cdot\|$ denotes the usual metric. Show that f must be a surjection.

DEFINITION 2.4.31. Recall that, in this chapter, F denotes either $\mathbb{R}$ or $\mathbb{C}$. Suppose that (X, d) is a metric space. Define $\mathcal{BC}(X, F)$ to be the subset of $\mathcal{B}(X, F)$ consisting of continuous functions from X to F. We take the metric on $\mathcal{BC}(X, F)$ to be the same as that on $\mathcal{B}(X, F)$.

THEOREM 2.4.32. *The space $\mathcal{BC}(X, F)$ is a complete metric space.*

Proof. Suppose that $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{BC}(X, F)$. Then by Theorem 2.4.14, $(f_n)_{n\in\mathbb{N}}$ converges to a function $f \in \mathcal{B}(X, F)$. All we need to show is that f is a continuous function. Now, given $\varepsilon > 0$, there exists N such that $\sup_{x\in X} |f_n(x) - f(x)| < \varepsilon/3$ whenever $n \geq N$. Fix $x_0 \in X$. Then, for any $x \in X$ and $n \geq N$,

$$\begin{aligned} |f(x) - f(x_0)| &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| \\ &< \varepsilon/3 + |f_n(x) - f_n(x_0)| + \varepsilon/3. \end{aligned}$$

Since f_n is continuous, we can choose $\delta > 0$ such that $|f_n(x) - f_n(x_0)| < \varepsilon/3$ whenever $d(x, x_0) < \delta$. But then $|f(x) - f(x_0)| < \varepsilon$ when $d(x, x_0) < \delta$, so f is continuous.

REMARK 2.4.33. So we have proved that the uniform limit of bounded functions is a bounded function and the uniform limit of bounded continuous functions is a bounded continuous function. We will find these facts very useful in doing analysis.

EXERCISE 2.4.34. Show that $\mathcal{BC}(X, F)$ is a subalgebra of $\mathcal{B}(X, F)$. That is, $\mathcal{BC}(X, F)$ is a vector subspace of $\mathcal{B}(X, F)$ which is closed under pointwise multiplication.

EXERCISE 2.4.35. In Remark 2.4.12, we saw an example in which the pointwise limit of a sequence of functions is continuous even though the convergence was not uniform. Now consider the sequence of functions $f_n : [0,1] \longrightarrow [0,1]$ where $f_n(x) = x^n$. Find the pointwise limit of the sequence $(f_n)_{n\in\mathbb{N}}$ and show that it is not continuous.

EXERCISE 2.4.36. Define a sequence of functions $f_n : (0,1) \longrightarrow \mathbb{R}$ by

$$f_n(x) = \begin{cases} \frac{1}{q^n} & \text{if } x = \frac{p}{q} \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

for $n \in \mathbb{N}$. Find the pointwise limit f of the sequence $(f_n)_{n\in\mathbb{N}}$ and show that $(f_n)_{n\in\mathbb{N}}$ converges to f uniformly.

There is an additional property of continuous functions which is important for future applications.

DEFINITION 2.4.37. Let (X,d) and (X',d') be metric spaces, and let f be a continuous function from X to X'. We say that f is *uniformly continuous* if, given $\varepsilon > 0$, there exists $\delta > 0$ such that, for any pair $x, y \in X$, we have $d'(f(x), f(y)) < \varepsilon$ whenever $d(x,y) < \delta$.

So, f is uniformly continuous if it is continuous at every point and, for a given $\varepsilon > 0$, we can find a corresponding δ that is independent of the point.

EXERCISE 2.4.38. Let $X = X' = \mathbb{R}$ with the usual metric.

(i) Show that a polynomial function $p(x)$ on $\mathbb{R}$ is uniformly continuous if and only if $\deg(p(x)) < 2$.
(ii) Show that $f(x) = \sin(x)$ is uniformly continuous on $\mathbb{R}$.

EXERCISE 2.4.39. Let $X = (0,\infty)$ and determine whether the following functions are uniformly continuous on X:

(i) $f(x) = 1/x$;
(ii) $f(x) = \sqrt{x}$;
(iii) $f(x) = \ln(x)$;
(iv) $f(x) = x\ln(x)$.

EXERCISE 2.4.40. Let $[a,b]$ be a closed bounded interval in $\mathbb{R}$ and let $f : [a,b] \longrightarrow \mathbb{R}$ be continuous. Show that f is uniformly continuous on $[a,b]$.

2.5. Absolute Continuity and Bounded Variation in $\mathbb{R}$

In the following discussion, we say that two intervals are *nonoverlapping* if they intersect at most at one endpoint.

DEFINITION 2.5.1. Let $[a,b]$ be a closed interval in $\mathbb{R}$. A function $f : [a,b] \longrightarrow \mathbb{R}$ is *absolutely continuous* if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that for any collection of pairwise nonoverlapping subintervals $[\alpha_1, \beta_1], \ldots,$

$[\alpha_n, \beta_n]$ of total length less than δ, we have

$$\sum_{i=1}^{n} |f(\beta_i) - f(\alpha_i)| < \varepsilon.$$

EXERCISE 2.5.2.

(i) Show that if $f : [a, b] \longrightarrow \mathbb{R}$ is absolutely continuous, then f is uniformly continuous, and hence continuous.
(ii) Find an example of a function $f : [a, b] \longrightarrow \mathbb{R}$ which is uniformly continuous but not absolutely continuous.

DEFINITION 2.5.3. Let $[a, b]$ be a closed interval in $\mathbb{R}$. A function $f : [a, b] \longrightarrow \mathbb{R}$ is of *bounded variation* if there exists a constant M such that, for every partition

$$P = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$$

of $[a, b]$, the *variation*

$$\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| < M.$$

The *total variation* of f on $[a, b]$ is the least upper bound of the set of variations associated to all partitions of $[a, b]$.

Recall that a real-valued function defined on an interval I in $\mathbb{R}$ is called *monotonic (or monotone) increasing* on I if $x < y$ implies $f(x) \leq f(y)$ for all $x, y \in I$. The function f is called *strictly monotonic (or strictly monotone) increasing* on I if $x < y$ implies $f(x) < f(y)$ for all $x, y \in I$. Similar definitions hold for decreasing functions.

EXERCISE 2.5.4.

(i) Let $f : [0, \pi] \longrightarrow \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x \cos(\pi/x), & x \neq 0, \\ 0, & x = 0. \end{cases}$$

Show that f is uniformly continuous on $[0, \pi]$ but is not of bounded variation.
(ii) Show that monotone functions are of bounded variation.
(iii) Show that any function of bounded variation can be written as the difference of two monotone functions.

THEOREM 2.5.5. *If $f : [a, b] \longrightarrow \mathbb{R}$ is absolutely continuous, then f is of bounded variation.*

Proof. By the absolute continuity of f, given $\varepsilon = 1$, there is a $\delta > 0$ such that

$$\sum_{i=1}^{m} |f(\beta_i) - f(\alpha_i)| < 1$$

whenever the nonoverlapping subintervals $[\alpha_i, \beta_i]$ satisfy $\sum(\beta_i - \alpha_i) < \delta$. Given a partition

$$P = \{a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b\}$$

of $[a, b]$, we consider the variation $\mathrm{var}(f, P)$ of f over P. If P' is a refinement of P, then it is clear that $\mathrm{var}(f, P') \geq \mathrm{var}(f, P)$. Pick N such that $\delta' = (b-a)/N < \delta$, and refine P to a partition P' by adding the points $a+\delta', a+2\delta', \ldots, a+(N-1)\delta'$. Then,

$$\mathrm{var}(f, P) \leq \mathrm{var}(f, P') = \sum_{i=0}^{N-1} \mathrm{var}(f, P' \cap [a + i\delta', a + (i+1)\delta']) < N$$

and hence f is of bounded variation.

At this point, we consider a monotonic function $c : [0, 1] \to [0, 1]$ which is continuous, and hence uniformly continuous, but is not absolutely continuous. This is probably not the example that you produced in Exercise 2.5.2.

Example 2.5.6. Consider the Cantor set produced in Exercise 1.6.26. Recall that, at the n-th stage of the construction of this set, we removed a union of 2^{n-1} open intervals from $[0, 1]$, each of length 3^{-n}. Thus, after n stages, a total of

$$1 + 2 + \cdots + 2^{n-1} = 2^n - 1$$

intervals are removed.

Define a sequence of continuous nondecreasing functions $c_n : [0, 1] \longrightarrow [0, 1]$ inductively as follows. Let $c_0(x) = x$ and, for $n \geq 0$, we set

$$c_{n+1}(x) = \begin{cases} \frac{1}{2}c_n(3x), & 0 \leq x \leq 1/3, \\ 1/2, & 1/3 \leq x \leq 2/3, \\ \frac{1}{2}c_n(3x-2) + \frac{1}{2}, & 2/3 \leq x \leq 1. \end{cases}$$

Exercise 2.5.7.

(i) Show that c_n is constant on each of the open intervals removed in the first n stages of the construction of the Cantor set and that moreover the value of c_n on the k-th such subinterval (ordered as they occur in $[0, 1]$) is $k/2^n$ for $0 < k < 2^n - 1$.

(ii) Deduce that if $m \geq n \geq 1$ and x is an element of the union of the intervals removed in the first n stages of the construction of the Cantor set, then $c_m(x) = c_n(x)$.

We claim that the sequence $(c_n)_{n\in\mathbb{N}}$ converges uniformly to a function $c : [0, 1] \longrightarrow [0, 1]$. Indeed, it is clear from the inductive definition that

$$\sup_{x\in[0,1]} |c_{n+1}(x) - c_n(x)| = \frac{1}{2} \sup_{x\in[0,1]} |c_n(x) - c_{n-1}(x)| \quad \text{for all} \quad n \geq 1.$$

It follows that

$$\sup_{x\in[0,1]} |c_{n+1}(x) - c_n(x)| = \frac{1}{2^n} \sup_{x\in[0,1]} |c_1(x) - c_0(x)| \quad \text{for all} \quad n \geq 0$$

and thus, for $m > n$,

$$\begin{aligned}\sup_{x\in[0,1]} |c_m(x) - c_n(x)| &\leq \sup_{x\in[0,1]} |c_m(x) - c_{m-1}(x)| \\ &\quad + \cdots + \sup_{x\in[0,1]} |c_{n+1}(x) - c_n(x)| \\ &= \left(\frac{1}{2^{m-1}} + \cdots + \frac{1}{2^n}\right) \sup_{x\in[0,1]} |c_1(x) - c_0(x)| \\ &\leq \frac{1}{2^{n-1}} \sup_{x\in[0,1]} |c_1(x) - c_0(x)|.\end{aligned}$$

It follows that the sequence $(c_n)_{n\in\mathbb{N}}$ is Cauchy and hence converges to a continuous function c on $[0,1]$ by Theorem 2.4.32. Clearly f is nondecreasing and has image in $[0,1]$ since this is true of each of the c_n. Moreover it is clear that $c(0) = 0$ and $c(1) = 1$, so by the Intermediate Value Theorem from elementary calculus (or see Corollary 2.6.34), $c : [0,1] \longrightarrow [0,1]$ is onto. Exercise 2.5.7 implies immediately that the image under c of the complement of the Cantor set is the set of rational numbers of the form $k/2^n$ for some $n \geq 1$ and some $0 < k < 2^n - 1$ and is, in particular, countable.

For each integer $n \geq 1$, the remaining intervals at the n-th stage of the construction of the Cantor set comprise a collection of 2^n disjoint closed intervals $[\alpha_i, \beta_i]$ of total length $(2/3)^n$. Moreover, by Exercise 2.5.7, we have

$$\sum_{i=1}^{2^n} |c(\beta_i) - c(\alpha_i)| = \sum_{i=1}^{2^n} \frac{i}{2^n} - \frac{i-1}{2^n} = 1.$$

It follows immediately from the definition that c cannot be absolutely continuous.

Recall that, given a nonempty bounded set B of real numbers, we often denote $\operatorname{lub} B$ by $\sup B$ and $\operatorname{glb} B$ by $\inf B$.

DEFINITION 2.5.8. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. Then, the *limit supremum* ($\limsup$) of $(x_n)_{n\in\mathbb{N}}$ is

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \sup_{m\geq n} \{x_m\}$$

and the *limit infimum* ($\liminf$) of $(x_n)_{n\in\mathbb{N}}$ is

$$\liminf_{n\to\infty} = \lim_{n\to\infty} \inf_{m\geq n} \{x_m\}.$$

Note that if $(x_n)_{n\in\mathbb{N}}$ is not bounded above, then $\limsup_{n\to\infty} x_n = +\infty$. Similarly, if $(x_n)_{n\in\mathbb{N}}$ is not bounded below, then $\liminf_{n\to\infty} x_n = -\infty$.

EXERCISE 2.5.9.

(i) Show that $\limsup_{n\to\infty} x_n$ is finite iff $(x_n)_{n\in\mathbb{N}}$ is bounded above.
(ii) Show that $\liminf_{n\to\infty} x_n$ is finite iff $(x_n)_{n\in\mathbb{N}}$ is bounded below.
(iii) Show that $(x_n)_{n\in\mathbb{N}}$ is convergent iff $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n$.

EXERCISE 2.5.10.

(i) Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence of real numbers. Show that $c = \limsup_{n\to\infty} x_n$ iff for any $\varepsilon > 0$ there are only a finite number of n such that $x_n > c + \varepsilon$ and an infinite number of n such that $x_n > c - \varepsilon$.
(ii) Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence of real numbers. Show that $c = \liminf_{n\to\infty} x_n$ iff for any $\varepsilon > 0$ there are only a finite number of n such that $x_n < c - \varepsilon$ and an infinite number of n such that $x_n < c + \varepsilon$.
(iii) Suppose that $(x_n)_{n\in\mathbb{N}}$ is a bounded sequence of distinct real numbers. Show that $\limsup_{n\to\infty} x_n$ is the least upper bound of the accumulation points of the set $\{x_n \mid n \in \mathbb{N}\}$ and that $\liminf_{n\to\infty} x_n$ is the greatest lower bound of the accumulation points of the set $\{x_n \mid n \in \mathbb{N}\}$.

We can extend the notions of lim sup and lim inf to sequences of real-valued functions. This idea will be important in Chapter 6 when we discuss measurable functions from a set X to the real numbers.

DEFINITION 2.5.11. Let X be a set and let $(f_i)_{i\in\mathbb{N}}$ be a sequence of real-valued functions on X. Then, the *supremum* of (f_i) is the extended real-valued function defined by

$$\bigvee_{i\in\mathbb{N}} f_i(x) = \sup_{i\in\mathbb{N}} f_i(x).$$

Here, by the term "extended" we are allowing the possibility that the value of this function is $+\infty$. Similarly, the *infimum* of $(f_i)_{i\in\mathbb{N}}$ is the extended real-valued function

$$\bigwedge_{i\in\mathbb{N}} f_i(x) = \inf_{i\in\mathbb{N}} f_i(x)$$

where we now allow the value $-\infty$.

The *limit supremum* of the sequence $(f_i)_{i\in\mathbb{N}}$ is defined as

$$\limsup_{i\to\infty} f_i(x) = \bigwedge_{j\in\mathbb{N}} \bigvee_{i=j}^{\infty} f_i(x).$$

Finally, the *limit infimum* of $(f_i)_{i\in\mathbb{N}}$ is

$$\liminf_{i\to\infty} f_i(x) = \bigvee_{j\in\mathbb{N}} \bigwedge_{i=j}^{\infty} f_i(x).$$

REMARK 2.5.12. The symbols $\vee$ and $\wedge$ are often called "cup" and "cap", respectively. They represent an extension of the standard "lattice operations" on functions:

$$\begin{aligned}(f \vee g)(x) &= \max(f(x), g(x)),\\ (f \wedge g)(x) &= \min(f(x), g(x)).\end{aligned}$$

2.6. Compactness, Completeness, and Connectedness

Now we consider a distinguished class of subsets of metric spaces.

DEFINITION 2.6.1. Let A be a nonempty subset of a metric space X. A family $\{U_j\}_{j\in J}$ of open subsets of X is called an *open covering* (or *open cover*) of A if

$$A \subseteq \bigcup_{j\in J} U_j.$$

If $\{U_j\}_{j\in J}$ is an open cover of A, we say that this cover has a *finite subcovering* if there is a finite subcollection $U_{j_1}, U_{j_2}, \ldots, U_{j_n}$ satisfying

$$A \subseteq \bigcup_{k=1}^{n} U_{j_k}.$$

EXAMPLES 2.6.2.

(i) Let $A = (0,1) \subseteq \mathbb{R}$ with the usual metric. For $j \in \mathbb{N}$, define $U_j = (\frac{1}{j}, 1)$. Then $(0,1) \subseteq \bigcup_{j\in\mathbb{N}} U_j$, but there is no finite subcover.

(ii) Let X be a discrete metric space. For any point $j \in X$, set $U_j = \{j\}$. Then $\{U_j\}_{j\in X}$ is an open cover of X which has a finite subcover iff X is a finite set.

(iii) We have seen in Chapter 1, Theorem 1.6.27, that if A is a closed and bounded set in $\mathbb{R}$ with the usual metric, then every open cover of A has a finite subcover.

Here is one of the big ideas in the theory of metric spaces.

DEFINITION 2.6.3. Let A be a subset of a metric space X. We say that A is *compact* if **every** open covering of A has a finite subcovering.

Recall that the Heine-Borel theorem in $\mathbb{R}$ states that a subset of $\mathbb{R}$ with the usual metric is compact if and only if it is closed and bounded. The statement of the Heine-Borel theorem is certainly not true in a general metric space. For example, take $\mathbb{R}$ with the discrete metric. Then $\mathbb{R}$ is closed and bounded in this metric. Take an open covering consisting of the individual points in $\mathbb{R}$. This covering does not have a finite subcovering.

For emphasis, we note that the definition insists that for **every** open covering, there must be a finite subcovering. For example, given any subset A of a metric space X, the set X is an open covering which is already a finite subcovering. So while this particular open covering has a finite subcover, this does not necessarily imply that other open coverings have finite subcoverings.

Hence, in a general metric space, the closed bounded sets are not necessarily compact. However, we do have one-half of the statement of the Heine-Borel theorem in general metric spaces.

THEOREM 2.6.4. *If a subset A of a metric space X is compact, then A is closed and bounded.*

Proof. Recall that a set in a metric space is bounded if and only if it is contained in a ball of finite radius with center at some point. If A is nonempty, take a point $a \in A$ and consider the open covering $\{B_n(a) \mid n \in \mathbb{N}\}$. Since A is compact, this cover has a finite subcovering, and in fact there is an integer N so that $A \subseteq B_N(a)$. Hence, A is bounded.

To prove that A is closed, we assume that x_0 is an accumulation point of A and prove that $x_0 \in A$. Suppose not. Then for each $a \in A$, let $r_a = d(a, x_0)/2$. But the collection $\{B_{r_a}(a) \mid a \in A\}$ is an open cover of A and hence has a finite subcover $\{B_{r_1}(a_1), B_{r_2}(a_2), \ldots, B_{r_n}(a_n)\}$. Let $r = \min\{r_1, r_2, \ldots, r_n\}$. Then $B_r(x_0) \cap B_{r_j}(a_j) = \varnothing$ for all j. Hence, $B_r(x_0) \cap A = \varnothing$, which contradicts the definition of accumulation point. Thus $x_0 \in A$.

COROLLARY 2.6.5. *If A is a compact set in a metric space X, then every infinite subset of A has an accumulation point in A.*

Proof. Suppose A is a compact set and that C is an infinite subset of A with no accumulation point in A. Then, for each $a \in A$, there is an open ball $B(a)$ centered at a such that $(B(a) \setminus \{a\}) \cap C = \varnothing$. The collection $\{B(a) \mid a \in A\}$ covers A. So, by compactness, we can extract a finite subcover, $\{B(a_1), \ldots, B(a_n)\}$. Thus, $C \subseteq A \subseteq B(a_1) \cup \cdots \cup B(a_n)$, and each $B(a_j)$ contains at most one element of C (at its center). This implies that C has at most n elements.

COROLLARY 2.6.6. *Let A be a compact set in a metric space. Then every infinite sequence in A has a subsequence that converges to a point in A.*

EXERCISE 2.6.7. Show that a closed subset of a compact set is compact.

EXERCISE 2.6.8. Prove that the Heine-Borel theorem holds in $\mathbb{R}^n$ and $\mathbb{C}^n$ with the usual metrics. (Hint: See the proof of Theorem 1.9.17.)

EXERCISE 2.6.9.

(i) Let $f : X \longrightarrow X'$ be a continuous map of metric spaces. Show that if $A \subseteq X$ is compact, then $f(A) \subseteq X'$ is compact.

(ii) Suppose that X is a compact metric space. Show that a continuous function $f : X \longrightarrow \mathbb{R}$ ($\mathbb{R}$ with the usual metric) is bounded.

(iii) Suppose that X is a compact metric space. Show that a continuous function $f : X \longrightarrow \mathbb{R}$ ($\mathbb{R}$ with the usual metric) attains a maximum and minimum value on X.

EXERCISE 2.6.10. Suppose X and X' are metric spaces with X compact.

(i) If $f : X \longrightarrow X'$ is continuous on X, show that f is uniformly continuous on X.

(ii) If $f : X \longrightarrow X'$ is a continuous bijection, show that f is a homeomorphism.

EXERCISE 2.6.11 **(Dini's theorem)**. Let X be a compact metric space. Suppose f and $(f_n)_{n\in\mathbb{N}}$ are real-valued continuous functions on X. Suppose that, for each $x \in X$, the sequence $(f_n(x))_{n\in\mathbb{N}}$ is a monotonic sequence converging to $f(x)$. Show that $(f_n)_{n\in\mathbb{N}}$ converges to f uniformly.

EXERCISE 2.6.12. Suppose that A and B are nonempty subsets of a metric space X. The *distance* between A and B is defined by

$$d(A, B) = \inf\{d(a, b) \mid a \in A, b \in B\}.$$

We say that $d(A, B)$ is *assumed* if there exists $a_0 \in A$ and $b_0 \in B$ such that $d(A, B) = d(a_0, b_0)$. Determine whether or not the distance between A and B is necessarily assumed in each of the following cases:

(i) A is closed and B is closed.
(ii) A is compact and B is closed.
(iii) A is compact and B is compact.
(iv) What happens in the above cases if we assume X is complete?

EXERCISE 2.6.13. Let X be a compact metric space and assume that $f : X \longrightarrow X$ is distance-preserving. That is, assume that $d(f(x_1), f(x_2)) = d(x_1, x_2)$ for all $x_1, x_2 \in X$. Show that f is surjective. Note: You might wish to use this result in solving Exercise 2.4.30.

At this point we introduce an alternate notion of compactness.

DEFINITION 2.6.14. A subset A of a metric space X is *sequentially compact* if every sequence in A has a subsequence that converges to an element of A.

EXERCISE 2.6.15. If X is a metric space and $A \subseteq X$, we say that A is *totally bounded* if, for any $\varepsilon > 0$, A can be covered by a finite number of balls of radius ε. Show that a sequentially compact metric space is totally bounded.

One of the most important facts about metric spaces is that compactness and sequential compactness are equivalent. We have just finished the proof that compactness implies sequential compactness. To prove the converse, we need the following lemma.

LEMMA 2.6.16. *Let X be a metric space. If $A \subseteq X$ has the property that every infinite subset of A has an accumulation point in X, then there exists a countable collection of open sets $\{U_i \mid i \in \mathbb{N}\}$ such that, if V is any open set in X and $x \in A \cap V$, then there is some U_i such that $x \in U_i \subseteq V$.*

Proof. We claim that, for each $n \in \mathbb{N}$, there is a finite set of points $x_1, \ldots, x_{N(n)}$ in A such that the set of open balls $B_{\frac{1}{n}}(x_1)$, $B_{\frac{1}{n}}(x_2), \ldots,$ $B_{\frac{1}{n}}(x_{N(n)})$ covers A. Otherwise, there exists $n \in \mathbb{N}$ such that no finite collection of balls of radius $\frac{1}{n}$ centered at points of A can cover A. For each $k \in \mathbb{N}$, assume A is infinite and define an infinite sequence of points of A inductively as follows. Take $y_1 \in A$. Then $B_{\frac{1}{n}}(y_1)$ does not cover A. So choose

$y_2 \in A \setminus B_{\frac{1}{n}}(y_1)$. Then $B_{\frac{1}{n}}(y_1) \cup B_{\frac{1}{n}}(y_2)$ does not cover A and $d(y_1, y_2) \geq \frac{1}{n}$. Assume $y_1, \ldots, y_k$ have been chosen so that $B_k = B_{\frac{1}{n}}(y_1) \cup \cdots \cup B_{\frac{1}{n}}(y_k)$ does not cover A and $d(y_i, y_j) \geq \frac{1}{n}$ for all $i \neq j$. Choose $y_{k+1} \in A \setminus B_k$. The infinite sequence $(y_k)_{k \in \mathbb{N}}$ does not have an accumulation point anywhere, which is a contradiction. Taking all these balls $\{B_{\frac{1}{n}}(x_j) \mid n \in \mathbb{N} \text{ and } 1 \leq j \leq N(n)\}$ gives the required collection.

EXERCISE 2.6.17. Verify that the above collection satisfies the conclusion of the lemma.

EXERCISE 2.6.18. Let X be a metric space. If $A \subseteq X$ has the property that every infinite subset of A has an accumulation point in A, show that for any open covering of A, there exists a countable subcovering.

Now comes a major theorem.

THEOREM 2.6.19. *In any metric space, a subset A is compact if and only if it is sequentially compact.*

Proof. We have already proved above that compactness implies sequential compactness. For the converse, suppose that $A \subseteq X$ is sequentially compact. Then any infinite subset of A contains a countable subset, which defines a sequence in A. By sequential compactness, this sequence has a subsequence that converges to a point $a \in A$. Since this point is clearly an accumulation point of A, we can apply Lemma 2.6.16 and Exercise 2.6.18 to conclude that, for any open cover $\mathcal{U}$ of A, we can find a countable subcover $\mathcal{U}'$.

From this open cover $\mathcal{U}' = \{U_j \mid j \in \mathbb{N}\}$, we wish to extract a finite subcover. Suppose that, for each n, the union $U_1 \cup \cdots \cup U_n$ does not cover A. Then, for each n, there exists $x_n \in A \setminus (U_1 \cup \cdots \cup U_n)$. This defines a sequence $(x_n)_{n \in \mathbb{N}}$ in A which by sequential compactness has a limit point $x \in A$. Since $\mathcal{U}'$ covers A, x must be contained in U_N for some N. But then U_N contains infinitely many elements of the sequence and hence contains some x_m with $m > N$. This is a contradiction.

EXERCISE 2.6.20.

(i) Show that a compact metric space is complete.
(ii) Show that a totally bounded complete metric space is compact.

Note that Theorem 2.6.19 gives another proof of the Heine-Borel theorem in $\mathbb{R}^n$ and $\mathbb{C}^n$.

EXERCISE 2.6.21. Prove this.

Compact sets in $\mathbb{R}^n$ with the usual metric have many interesting properties, some of which are illustrated in the following exercises.

EXERCISE 2.6.22. Let B be a compact convex subset of $\mathbb{R}^n$ with the usual metric. Define the *nearest point function* $p : {}^cB \longrightarrow B$ as follows: for

$x \in {}^cB$ we set $p(x)$ to be the closest point to x that lies in B. Show the following:

(i) the function $p(x)$ is well-defined;
(ii) the point $p(x)$ lies in the boundary of B;
(iii) the function $p(x)$ is surjective onto the boundary of B.

In the next exercise, we continue with the terminology of the preceding exercise. Define the supporting hyperplane at $p(x)$ to be the hyperplane through $p(x)$ orthogonal to the vector $p(x) - x$. Define the supporting half-space at $p(x)$ to be the set

$$H_{p(x)} = \{y \in \mathbb{R}^n \mid (y - p(x)) \cdot (p(x) - x) \geq 0\}.$$

Here, "$\cdot$" denotes the usual dot product in $\mathbb{R}^n$.

EXERCISE 2.6.23.

(i) Show that, for each $x \in {}^cB$, the set B is a subset of $H_{p(x)}$.
(ii) Show that $B = \bigcap_{y \in \partial B} H_y$.
(iii) Does the above process work when B is a closed convex unbounded subset of $\mathbb{R}^n$ with the usual metric?

Here are a few more interesting facts and ideas about metric spaces. The first involves the notion of separability.

DEFINITION 2.6.24. Let (X, d) be a metric space. A subset $A \subseteq X$ is said to be *dense* in X if $\overline{A} = X$.

EXERCISE 2.6.25. It is clear that X is dense in X. Is it possible that the only subset of X which is dense in X is X itself?

DEFINITION 2.6.26. Let (X, d) be a metric space. We say that X is *separable* if there exists a countable subset of X which is dense in X.

EXAMPLE 2.6.27. The spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ with the usual metric are separable. As a countable dense subset, we can take the collection of all points in $\mathbb{R}^n$ whose coordinates are rational numbers, or the set of all points in $\mathbb{C}^n$ whose coordinates have the property that their real and imaginary parts are rational numbers.

THEOREM 2.6.28. *If (X, d) is a compact metric space, then X is separable.*

Proof. For each $n \in \mathbb{N}$, consider the collection of open balls $\{B_{1/n}(x) \mid x \in X\}$. This is an open covering of X, and hence, there is a finite subcovering $\mathcal{U}_n$. Take the union over $n \in \mathbb{N}$ of the centers of the balls in $\mathcal{U}_n$. This is a countable collection of points in X which is obviously dense.

EXERCISE 2.6.29. Suppose X and X' are metric spaces with X separable. Let $f : X \longrightarrow X'$ be a continuous surjection. Show that X' is separable.

As shown in Example 2.6.27, separable metric spaces do not have to be compact. Many of the important metric spaces which occur in analysis are separable, but there are some very important examples of nonseparable metric spaces. This will be discussed further in Chapter 4.

Another important idea in metric spaces is connectedness. It has a funny definition because we begin by defining a disconnected set.

DEFINITION 2.6.30. Let X be a metric space and let $A \subseteq X$. We say that A is *not connected* (or *disconnected*) if there exist open sets $U, V \subseteq X$ such that

(a) $U \cap A \neq \varnothing$ and $V \cap A \neq \varnothing$,
(b) $(U \cap A) \cap (V \cap A) = \varnothing$,
(c) $A = (U \cap A) \cup (V \cap A)$.

We say that A is *disconnected* by the open sets U and V.

DEFINITION 2.6.31. Let X be a metric space and let $A \subseteq X$. We say A is *connected* if A is not disconnected.

EXERCISE 2.6.32.

(i) Show that a subset A of $\mathbb{R}$ in the usual metric is connected iff A is an interval.
(ii) Show that a convex subset of $\mathbb{R}^n$ with the usual metric is a connected set.

The basic theorem about connected sets is the following.

THEOREM 2.6.33. *Let X, X' be metric spaces and let $f : X \longrightarrow X'$ be a continuous function. If A is a connected subset of X, then $f(A)$ is a connected subset of X'. That is, the continuous image of a connected set is connected.*

Proof. Let U and V be open sets in X', and assume that U and V disconnect $f(A)$. Then, $f^{-1}(U)$ and $f^{-1}(V)$ are open sets in X which disconnect A.

COROLLARY 2.6.34 **(Intermediate Value Theorem)**. *Let X be a metric space, and take $\mathbb{R}$ with the usual metric. Let $f : X \longrightarrow \mathbb{R}$ be a continuous function. Let A be a connected subset of X and let $I = f(A)$. Then I is an interval in $\mathbb{R}$.*

EXERCISE 2.6.35. Use the corollary to show the following.

(i) Take $\mathbb{R}$ with the usual metric, and let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be given by $f(x) = x^n$ for $n \in \mathbb{N}$. If b is a positive real number, show that there exists a unique positive real number a such that $a^n = b$.
(ii) Show that any polynomial of odd degree has a root in $\mathbb{R}$.

EXERCISE 2.6.36.

(i) Show that an open ball in $\mathbb{R}^n$ or $\mathbb{C}^n$ with the usual metric is a connected set.
(ii) Show that the closed ball in $\mathbb{R}^n$ or $\mathbb{C}^n$ with the usual metric is a connected set.
(iii) Show that $GL_2(\mathbb{R})$ with the metric inherited from $M_2(\mathbb{R})$ as in Exercise 2.3.31 is not a connected set. (Hint: Use the fact that the determinant is a continuous function.)
(iv) Show that $GL_2(\mathbb{C})$ with the metric inherited from $M_2(\mathbb{C})$ is a connected set.

If a metric space X is not connected, then it can be decomposed into subsets called connected components.

DEFINITION 2.6.37. If X is a metric space and x_0 is in X, then the *connected component of* x_0 *in* X is the union of the connected sets that contain x_0.

EXERCISE 2.6.38.

(i) Let X be a metric space and take $x_0 \in X$. Show that the connected component of x_0 is a connected set in X.
(ii) Show that if A is a connected subset of a metric space, then $\overline{A}$ is connected.

EXAMPLES 2.6.39.

(i) Let $X = \mathbb{R}^\times$ be the set of nonzero real numbers with the usual metric. This metric space has two connected components, namely, the positive real numbers and the negative real numbers.
(ii) The connected components of $GL_2(\mathbb{R})$ with the usual metric are $GL_2^+(\mathbb{R}) = \{x \in GL_2(\mathbb{R}) \mid \det x > 0\}$ and $GL_2^-(\mathbb{R}) = \{x \in GL_2(\mathbb{R}) \mid \det x < 0\}$.

DEFINITION 2.6.40. A metric space X is *totally disconnected* if the connected component of each point is the point itself.

EXAMPLE 2.6.41. A discrete metric space X is totally disconnected.

EXERCISE 2.6.42.

(i) Find an example of a metric space which is totally disconnected but not discrete.
(ii) Find an example of a complete metric space which is totally disconnected but not discrete.

Obviously, complete metric spaces play a special role among all metric spaces. We now present a procedure through which any metric space can be embedded as a dense subset of a complete metric space.

THEOREM 2.6.43. *Let (X, d) be a metric space. Then there exists a complete metric space $(\tilde{X}, \tilde{d})$ and an injection $\phi : X \longrightarrow \tilde{X}$ such that*

(i) *$\phi : X \longrightarrow \phi(X)$ is an isometry,*
(ii) *$\phi(X)$ is dense in $\tilde{X}$.*

Proof. Consider the set X' of all Cauchy sequences in X. We define an equivalence relation on X' by saying that $(x_n)_{n\in\mathbb{N}}$ is equivalent to $(y_n)_{n\in\mathbb{N}}$ if $\lim_{n\longrightarrow\infty} d(x_n, y_n) = 0$.

EXERCISE 2.6.44. Prove that this is an equivalence relation.

Let $\tilde{X}$ be the set of equivalence classes. We first define a metric on $\tilde{X}$. Let $\{(x_n)_{n\in\mathbb{N}}\}$ and $\{(x'_n)_{n\in\mathbb{N}}\}$ be elements of $\tilde{X}$. We note that $(d(x_n, x'_n))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. This follows from the fact that $|d(x_n, x'_n) - d(x_m, x'_m)| \leq d(x_n, x_m) + d(x'_n, x'_m)$. We set

$$\tilde{d}(\{(x_n)_{n\in\mathbb{N}}\}, \{(x'_n)_{n\in\mathbb{N}}\}) = \lim_{n\longrightarrow\infty} d(x_n, x'_n).$$

This limit exists because the real numbers are known to be complete.

EXERCISE 2.6.45. Show that $\tilde{d}$ is well-defined.

Now define $\phi : X \longrightarrow \tilde{X}$ by $\phi(x) = \{(x_k)_{k\in\mathbb{N}}\}$ where $x_k = x$ for all $k \in \mathbb{N}$. It is clear that ϕ is an isometry from X to $\phi(X)$. There are two things left to do. First, show that $\phi(X)$ is dense in $\tilde{X}$, and second, show that $(\tilde{X}, \tilde{d})$ is complete.

Let $\tilde{x} = \{(x_n)_{n\in\mathbb{N}}\} \in \tilde{X}$. Pick $\varepsilon > 0$. Since the sequence $(x_n)_{n\in\mathbb{N}}$ is Cauchy in X, there exists an integer N such that $d(x_N, x_m) < \varepsilon$ if $m \geq N$. Now consider the class of the constant sequence $(x_N) = \phi(x_N)$. Then $\tilde{d}(\tilde{x}, \phi(x_N)) = \lim_{n\to\infty} d(x_n, x_N) \leq \varepsilon$ and hence $\phi(X)$ is dense in $\tilde{X}$.

To show that $\tilde{X}$ is complete, take a Cauchy sequence $(\tilde{y}_n)$ in $\tilde{X}$. Remember, each $\tilde{y}_n$ is an equivalence class of Cauchy sequences in X. For each $n \in \mathbb{N}$, by density, choose $\tilde{z}_n \in \phi(X)$ such that $\tilde{d}(\tilde{y}_n, \tilde{z}_n) < \frac{1}{n}$. Then

$$\tilde{d}(\tilde{z}_n, \tilde{z}_m) \leq \tilde{d}(\tilde{z}_n, \tilde{y}_n) + \tilde{d}(\tilde{y}_n, \tilde{y}_m) + \tilde{d}(\tilde{y}_m, \tilde{z}_m) < \frac{1}{n} + \tilde{d}(\tilde{y}_n, \tilde{y}_m) + \frac{1}{m}.$$

This implies that $(\tilde{z}_n)_{n\in\mathbb{N}}$ is Cauchy in $\tilde{X}$. Let $x_n = \phi^{-1}(\tilde{z}_n)$. Then, since ϕ is an isometry, $(x_n)_{n\in\mathbb{N}}$ is Cauchy in X. Let $\tilde{y}$ be the element of $\tilde{X}$ defined by the equivalence class of this Cauchy sequence; that is, $\tilde{y} = \{(x_n)_{n\in\mathbb{N}}\}$. Then,

$$\tilde{d}(\tilde{y}_n, \tilde{y}) \leq \tilde{d}(\tilde{y}_n, \tilde{z}_n) + \tilde{d}(\tilde{z}_n, \tilde{y}) < \frac{1}{n} + \tilde{d}(\tilde{z}_n, \tilde{y}).$$

Observe that $\tilde{d}(\tilde{z}_n, \tilde{y}) = \lim_{k\to\infty} d(x_n, x_k)$. Since $(x_n)_{n\in\mathbb{N}}$ is Cauchy in X, for n and k large, $d(x_n, x_k)$ can be made arbitrarily small.

DEFINITION 2.6.46. The metric space $\tilde{X}$ in the above theorem is called the *completion* of X.

EXERCISE 2.6.47. If (X, d) is already a complete metric space, show that (X, d) and $(\tilde{X}, \tilde{d})$ are isometric.

EXERCISE 2.6.48. Prove that $(\tilde{X}, \tilde{d})$ is unique up to isometry. That is, if (X', d') is a complete metric space such that X is isometric to a dense subset of X', then $(\tilde{X}, \tilde{d})$ and (X', d') are isometric.

REMARK 2.6.49. One might ask at this point, "Why did we write Chapter 1 at all?" Why not just take the rational numbers with the usual metric and complete them by the above process to get the real numbers? Sorry folks, but in the proof of the above theorem, we used the fact that the real numbers are complete. In Chapter 3, we will have a simple, yet significant, example of the completion of a metric space, namely, the p-adic completion of $\mathbb{Q}$ relative to a prime p. This emphasizes the fact that while $\mathbb{R}$ is the most familiar example of a completion of $\mathbb{Q}$ with respect to a metric, there are in fact infinitely many other completions of $\mathbb{Q}$.

EXERCISE 2.6.50. Let (X, d) be a metric space, and for any $x, y \in X$, let $d'(x, y) = \frac{d(x,y)}{1+d(x,y)}$.

(i) Show that d' defines a metric on X.
(ii) Show that U is open in (X, d) if and only if U is open in (X, d').
(iii) If a set A is compact in (X, d), is A necessarily compact in (X, d')?
(iv) If (X, d') is complete, is (X, d) necessarily complete?

This exercise is intended to illustrate that, without additional structure, metric spaces can be twisted, expanded, or shrunken without disturbing the open sets too badly. This possibility will be corrected when we study normed linear spaces in Chapter 4.

2.7. Independent Projects

2.7.1. General Point Set Topology.

2.7.1.1. Basic Notions

DEFINITION 2.7.1. Given a set X and $\mathcal{T} \subseteq \mathscr{P}(X)$, we say that $\mathcal{T}$ is a *topology* on X provided that:

(a) the sets $\varnothing$ and X are in $\mathcal{T}$;
(b) if $\{U_i\}_{i\in I} \subseteq \mathcal{T}$, then $\bigcup_{i\in I} U_i \in \mathcal{T}$;
(c) if $\{U_i\}_{i=1}^n \subseteq \mathcal{T}$, then $\bigcap_{i=1}^n U_i \in \mathcal{T}$.

The pair $(X, \mathcal{T})$ is commonly referred to as a *topological space*. When the topology is clear or already given, the symbol $\mathcal{T}$ is often suppressed in the notation.

EXERCISE 2.7.2.

(i) If X is a set and $\alpha = \{\varnothing, X\}$, show that α is a topology on X known as the *indiscrete topology.*
(ii) If X is a set and $\omega = \mathscr{P}(X)$, show that ω is a topology on X known as the *discrete topology.*
(iii) If X is a set and $\mathcal{T}$ is a topology on X, show that $\alpha \subseteq \mathcal{T} \subseteq \omega$ with α and ω as above.

EXERCISE 2.7.3. Let X be a metric space, and let $\mathcal{T}$ be the collection of all open sets. Show that $\mathcal{T}$ is a topology on X. The topology on $\mathbb{R}^n$ given by the usual metric is often called the *usual topology.*

EXERCISE 2.7.4. Let $X = \mathbb{R}$. Let $\mathcal{T} = \{\text{all sets with finite complement}\} \cup \{\varnothing\}$. Show that $\mathcal{T}$ is a topology on X. This is called the *finite complement topology*.

DEFINITION 2.7.5. Let $(X, \mathcal{T})$ be a topological space. A subset U of X is *open* relative to $\mathcal{T}$ if $U \in \mathcal{T}$. A subset V of X is *closed* relative to $\mathcal{T}$ if $X \setminus V \in \mathcal{T}$. We say that N is a *neighborhood* of $x \in X$ if there exists $U \in \mathcal{T}$ such that $x \in U \subseteq N$.

EXERCISE 2.7.6. Show that a subset U of X is open iff, for all $x \in U$, there is a neighborhood of x contained in U.

PROPOSITION 2.7.7. *The closed sets of a topological space X satisfy the following properties:*

(i) *the sets $\varnothing, X$ are closed;*
(ii) *if $\{V_i\}_{i \in I}$ is a collection of closed sets, then $\bigcap_{i \in I} V_i$ is closed;*
(iii) *if $\{V_i\}_{i=1}^n$ is a finite collection of closed sets, then $\bigcup_{i=1}^n V_i$ is closed.*

Proof. This is immediate from DeMorgan's laws.

DEFINITION 2.7.8. Given a topology $\mathcal{T}$ on X and a set $E \subseteq X$, a point $x \in X$ is called an *accumulation point* of E if, for all neighborhoods N of x, we have $(N \cap E) \setminus \{x\} \neq \varnothing$.

EXERCISE 2.7.9.

(i) Show that a set E is closed iff it contains all of its accumulation points.
(ii) Show that the union of a set with its accumulation points is closed.
(iii) Show that the intersection over all closed sets containing E is equal to the union of E with its accumulation points.

REMARK 2.7.10. The theory of closure, boundary, and interior operations can now be carried out for topological spaces in the same manner as it was for metric spaces.

EXERCISE 2.7.11. Formulate a definition of density in a general topological space.

We now turn to the notion of continuous functions in general topological spaces.

DEFINITION 2.7.12. If $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ are topological spaces, we say that a map $f : X \longrightarrow Y$ is *continuous* if, for all $V \in \mathcal{T}_Y$, the set $f^{-1}(V) \in \mathcal{T}_X$. In other words, the inverse image of every open set in Y is an open set in X. In addition, if $f : X \longrightarrow Y$ is a continuous bijection such that f^{-1} is also continuous, then we say that f is a *homeomorphism.*

REMARK 2.7.13. Recall that in metric spaces the definition of continuity given above is equivalent to the δ-ε definition.

PROPOSITION 2.7.14. *If X, Y, and Z are topological spaces and $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$ are continuous functions, then $g \circ f : X \longrightarrow Z$ is a continuous function.*

Proof. Let U be an open set in Z. Because g is continuous, the set $g^{-1}(U)$ is open in Y, and because f is continuous, the set $f^{-1}(g^{-1}(U)) = (g \circ f)^{-1}(U)$ is open in X.

DEFINITION 2.7.15. If X and Y are topological spaces, a map $f : X \longrightarrow Y$ is an *open map* if, for every open set U contained in X, the set $f(U)$ is open in Y.

EXERCISE 2.7.16.

(i) Show that a homeomorphism is an open map.
(ii) Find topological spaces X, Y and an open map $f : X \longrightarrow Y$ such that there exists a closed set V in X with $f(V)$ not closed in Y.
(iii) Let X and Y be metric spaces. Suppose $f : X \longrightarrow Y$ has the property that, for each ball $B \subseteq X$ centered at $x \in X$, we have that $f(B)$ contains a ball centered at $f(x)$. Show that f is an open mapping.

DEFINITION 2.7.17. Given a topological space $(X, \mathcal{T})$, a *base* (or *basis*) for the topology $\mathcal{T}$ on X is a collection $\mathcal{B} \subseteq \mathcal{T}$ such that every element of $\mathcal{T}$ is a union of elements of $\mathcal{B}$.

DEFINITION 2.7.18. Let X be a set, and let $\mathcal{C}$ be a collection of subsets of X that includes $\varnothing$ and X. The *topology generated by* $\mathcal{C}$ is the intersection of all the topologies on X that contain $\mathcal{C}$.

EXERCISE 2.7.19. Let $(X, \mathcal{T})$ be a topological space, and let $\mathcal{B}$ be a base for the topology. Show that $\mathcal{T}$ is the topology generated by $\mathcal{B}$.

EXERCISE 2.7.20. Given a collection $\mathcal{C}$ of subsets of a set X, show that arbitrary unions and finite intersections of elements of $\mathcal{C}$ are in the topology generated by $\mathcal{C}$. Is it possible that the topology generated by $\mathcal{C}$ contains more elements than these?

THEOREM 2.7.21. *Let $(X, \mathcal{T})$ be a topological space with a countable base $\mathcal{B}$. Let $E \subseteq X$ be an uncountable set. Then there exists $x \in E$ such that x is an accumulation point of E.*

Proof. Suppose that no point in E is an accumulation point of E. Then, given any $x \in E$, there exists an open set U_x so that $U_x \cap E = \{x\}$. As $\mathcal{B}$ is a basis, there exists $B_x \in \mathcal{B}$ such that $x \in B_x \subseteq U_x$. This implies $B_x \cap E = \{x\}$. Note that if $x, y \in E$ and $x \neq y$, then $B_x \neq B_y$. It follows that there is a one-to-one correspondence from E into $\mathcal{B}$ given by $x \mapsto B_x$. This contradicts the uncountability of E.

DEFINITION 2.7.22. A topological space $(X, \mathcal{T})$ is *separable* if there is a countable dense subset in X, that is, there is a countable subset $A \subseteq X$ such that $\overline{A} = X$.

EXERCISE 2.7.23.

(i) Show that a topological space $(X, \mathcal{T})$ that has a countable base is separable.
(ii) Prove or disprove: a topological space $(X, \mathcal{T})$ which is separable has a countable base.
(iii) Let X and Y be topological spaces, and suppose that there exists a continuous surjection $f : X \longrightarrow Y$. Prove or disprove: if X is separable, then Y is separable.

2.7.1.2. Separation Properties

Although metric spaces are the most common spaces in analysis, there are strong mathematical reasons for considering different topologies that occur in other parts of mathematics. A particular case of such differences is illustrated by separation properties. Here are some of the most common.

DEFINITION 2.7.24 **(Separation properties)**. Let $(X, \mathcal{T})$ be a topological space.

(a) The space $(X, \mathcal{T})$ is T_0 if for every $x, y \in X$, there is a neighborhood N of one of the points that does not contain the other point.
(b) The space $(X, \mathcal{T})$ is T_1 if points are closed.
(c) The space $(X, \mathcal{T})$ is T_2 (commonly called *Hausdorff*) if, for every $x, y \in X$, there are open sets U, V so that $x \in U$, $y \in V$, and $U \cap V = \varnothing$.
(d) The space $(X, \mathcal{T})$ is T_3 (commonly called *regular*) if it is T_1, and given any $x \in X$ and a neighborhood N of x, there is a closed neighborhood V of x such that $V \subseteq N$.
(e) The space $(X, \mathcal{T})$ is T_4 (commonly called *normal*) if it is T_1, and given two closed sets A and B with $A \cap B = \varnothing$, there exist open sets U and V such that $A \subseteq U$, $B \subseteq V$, and $U \cap V = \varnothing$.

EXERCISE 2.7.25. Show that metric spaces satisfy all of the above separation properties.

EXERCISE 2.7.26.

(i) Show that if $(X, \mathcal{T})$ is T_i, then it is T_{i-1} for $i = 1, 2, 3, 4$.
(ii) Let $X = \mathbb{R} \cup \{\infty\}$, and let $\mathcal{T} = \{E \subseteq X \mid {}^cE \text{ is finite and } \infty \in E\} \cup \{\varnothing\}$. Show that $(X, \mathcal{T})$ is T_0 but not T_1.
(iii) Let $X = \mathbb{R}$ with the finite complement topology. Show that X is T_1 but not T_2.
(iv) Find an example of a topological space that is T_2 but not T_3.
(v) Find an example of a topological space that is T_3 but not T_4.

The following lemma demonstrates the type of theorem that follows from assuming certain separation properties in a topological space.

LEMMA 2.7.27 **(Urysohn's lemma)**. *If A and B are disjoint, closed subsets of a normal topological space X, then there is a continuous function $f : X \longrightarrow [0, 1]$ such that $f(A) = \{0\}$ and $f(B) = \{1\}$.*

Proof. Let $D = \{p2^{-q} \mid p, q \in \mathbb{N}\} \cup \{0\}$. We define $F : D \longrightarrow \mathscr{P}(X)$ and then use it to get the desired function. For $t \in D$ with $t > 1$ let $F(t) = X$, and let $F(1) = X \setminus B$. Using the assumption that X is normal, we pick $F(0)$ to be some open set containing A such that $\overline{F(0)} \cap B = \varnothing$. For $t \in D$ with $0 < t < 1$, we write $t = (2m+1)2^{-n}$ and choose, inductively on n, $F(t)$ to be an open set containing $\overline{F(2m \cdot 2^{-n})}$ and such that $\overline{F(t)} \subseteq F((2m+2) \cdot 2^{-n})$. This is possible because X is normal. Now we define $f : X \longrightarrow [0, 1]$. Let $f(x) = \inf\{t \mid x \in F(t)\}$. The function is 0 on A because $A \subseteq F(t)$ for all $t \in D$. The function f is 1 on B because $F(t) \subseteq X \setminus B$ for all $t \leq 1$ and $F(t) = X$ for $t > 1$.

Now we show that f is continuous. Observe that all sets of the form $\{t \mid t < s\}$ and $\{t \mid t > s\}$ generate the topology on $[0, 1]$, and it suffices to check that the inverse images of these sets are open. Notice that

$$\{x \mid f(x) < s\} = \bigcup_{\{j \in D, j < s\}} F(j).$$

For each $j \in D$, the set $F(j)$ is open and the arbitrary unions of open sets are open. Likewise,

$$\{x \mid f(x) \leq s\} = \bigcap_{\{j \in D, j > s\}} \overline{F(j)}.$$

This is an intersection of closed sets and hence is closed. Thus, $f^{-1}(\{t \mid t > s\})$ is open. ☠

ADVENTURE 2.7.28. Let $(X, \mathcal{T})$ be a regular topological space with a countable basis for $\mathcal{T}$. Show that there exists a metric on X that gives the topology $\mathcal{T}$.

This is known as *Urysohn's Metrization Theorem*. It will take some work.

2.7.1.3. New Topologies from Old

DEFINITION 2.7.29 **(Relative topology)**. Let $(X, \mathcal{T})$ be a topological space, and let $Y \subseteq X$. The *relative topology* on Y is the collection $\{E \cap Y \mid E \in \mathcal{T}\}$.

The relative topology on Y is sometimes called the *subspace topology.*

EXERCISE 2.7.30.

(i) Let $(X, \mathcal{T})$ be a topological space, and let $Y \subseteq X$ have the relative topology. Show that the inclusion map $i : Y \longrightarrow X$ is continuous.
(ii) Let X, Y be as above. Determine which of the separation properties, $T_0, \ldots, T_4$, are inherited by the relative topology on Y.

DEFINITION 2.7.31. Let $(X_1, \mathcal{T}_1)$ and $(X_2, \mathcal{T}_2)$ be two topological spaces. Let $X = X_1 \times X_2$ be the Cartesian product. Then the *product topology* on X is the topology generated by the collection $\mathcal{C} = \{E_1 \times E_2 \mid E_1 \in \mathcal{T}_1 \text{ and } E_2 \in \mathcal{T}_2\}$.

It is clear that the above definition can be extended to the product of a finite number of topological spaces.

EXAMPLE 2.7.32.

(i) The spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ are standard examples of product spaces.
(ii) The space $\mathbb{C}$ may be regarded topologically as $\mathbb{R} \times \mathbb{R}$.

DEFINITION 2.7.33. Let $\{(X_i, \mathcal{T}_i)\}_{i \in I}$ be a collection of topological spaces, and let $X = \prod_{i \in I} X_i$ be the Cartesian product. Then the *product topology* on X is the topology generated by the collection

$$\mathcal{C} = \Big\{ \prod_{i \in I} E_i \mid E_i \in \mathcal{T}_i \text{ for all } i \in I \text{ and } E_i = X_i \text{ for all but finitely many } i \in I \Big\}.$$

EXERCISE 2.7.34.

(i) For each $i \in \mathbb{N}$, let X_i be the closed interval $[0, 1]$ with the usual topology. Find the open sets in $X = \prod_{i \in \mathbb{N}} X_i$.
(ii) Let $A = \{x \in X \mid x_1 \in [0, 1] \text{ and } x_i = 0 \text{ for } i \geq 2\}$. Is the relative topology on A the usual topology on $[0, 1]$?

EXERCISE 2.7.35. If the index set is infinite, is there a reason why we do not define the product topology to be generated by sets of the form $\prod_{i \in I} E_i$ where each $E_i \in \mathcal{T}_i$?

EXERCISE 2.7.36. Let $\{(X_i, \mathcal{T}_i)\}_{i \in I}$ be a collection of topological spaces and let $X = \prod_{i \in I} X_i$ with the product topology.

(i) For each $i \in I$, let $p_i : X \longrightarrow X_i$ be the projection, defined by $p_i(x) = x_i$. Show that p_i is continuous.
(ii) Let Y be a topological space and let $f : Y \longrightarrow X$. Show that f is continuous if and only if $p_i \circ f$ is continuous for all $i \in I$.

DEFINITION 2.7.37. Let $(X, \mathcal{T})$ be a topological space, and let Y be a set. If $f : X \longrightarrow Y$ is a function, we define the *quotient topology* $\mathcal{T}'$ on Y as follows. Given $E \subseteq Y$, we say that $E \in \mathcal{T}'$ if $f^{-1}(E) \in \mathcal{T}$.

EXERCISE 2.7.38. Show that with this definition, $(Y, \mathcal{T}')$ is a topological space.

One of the most natural examples of the quotient topology occurs when $X = G$, a topological group, and H is a subgroup of G. We then take Y to be the coset space G/H and give it the quotient topology associated with the function $g \mapsto gH$.

EXERCISE 2.7.39. Let $G = \mathbb{R}$ with the usual topology, and let $H = \mathbb{Z}$. Show that G/H with the quotient topology is homeomorphic to S^1.

EXERCISE 2.7.40. Let X be a topological space and let Y be a space with the quotient topology given by the function $f : X \longrightarrow Y$. If Z is a topological space, show that $g : Y \longrightarrow Z$ is continuous iff $g \circ f$ is continuous.

2.7.1.4. Compactness

DEFINITION 2.7.41. A subset K of a topological space X is *compact* if every cover of K by open sets admits a finite subcover. We say that a space X is *locally compact* if every $x \in X$ has a compact neighborhood in X.

EXERCISE 2.7.42.

(i) Show that every closed subset of a compact space is compact.
(ii) Show that every closed subset of a locally compact space is locally compact.

EXERCISE 2.7.43. If X is Hausdorff, then any compact set $K \subseteq X$ is closed.

EXERCISE 2.7.44. If $f : X \longrightarrow Y$ is a continuous map, then for any compact set $K \subseteq X$, the set $f(K)$ is compact in Y.

EXERCISE 2.7.45. Show that if X is compact and Y is Hausdorff, then every bijective continuous function $f : X \longrightarrow Y$ is a homeomorphism.

ADVENTURE 2.7.46. Prove Tychonoff's theorem, which states the following: if $\{(X_i, \mathcal{T}_i)\}_{i \in I}$ is a collection of compact topological spaces, then $X = \prod_{i \in I} X_i$ with the product topology is a compact topological space.

2.7.2. The Fundamental Theorem of Algebra.

EXERCISE 2.7.47. Let P be a polynomial of positive degree with coefficients in $\mathbb{C}$. Show that there exists $z_0 \in \mathbb{C}$ such that $|P(z_0)| \leq |P(z)|$ for all $z \in \mathbb{C}$. Then show that, by considering the polynomial $P(z + z_0)$, we may assume $z_0 = 0$.

THEOREM 2.7.48 **(The Fundamental Theorem of Algebra)**. *The field $\mathbb{C}$ is algebraically closed; that is, any nonconstant polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$.*

Proof. Let $P \in \mathbb{C}[z]$ be a polynomial of positive degree. By Exercise 2.7.47, we may assume $P(z)$ has a minimum at 0. There exist $n \geq 1$ and $a, b \in \mathbb{C}$ with $b \neq 0$ so that

$$P(z) = a + bz^n + z^{n+1}Q(z),$$

where $Q \in \mathbb{C}[z]$. Suppose that $P(0) = a \neq 0$, and choose an n-th root w of $-a/b$ in $\mathbb{C}$.

By continuity, there exists t with $0 < t < 1$ so that $t|w^{n+1}Q(tw)| < |a|$. Now, we have

$$\begin{aligned} P(tw) &= a + b(tw)^n + (tw)^{n+1}Q(tw) \\ &= (1 - t^n)a + (tw)^{n+1}Q(tw) \end{aligned}$$

because $bw^n = -a$. Hence,

$$\begin{aligned} |P(tw)| &\leq (1 - t^n)|a| + t^{n+1}|w^{n+1}Q(tw)| \\ &< (1 - t^n)|a| + t^n|a| = |a| = |P(0)|. \end{aligned}$$

This is a contradiction, and hence we must have $P(0) = a = 0$.

EXERCISE 2.7.49. Find ten other proofs of the Fundamental Theorem of Algebra.

Starting with any field F, we wish to define an algebraic closure of F. We first define an algebraic extension of F.

DEFINITION 2.7.50. Let F be a field and let E be a field containing F as a subfield. We say that E is an *algebraic extension* of F if, given $\alpha \in E$, there exists a nonzero polynomial $p(x) \in F[x]$ such that $p(\alpha) = 0$.

DEFINITION 2.7.51. Let F be a field. A field E containing F is an *algebraic closure* of F if E is an algebraic extension of F and E is algebraically closed.

The following sequence of statements leads to the existence and uniqueness, up to isomorphism, of an algebraic closure of F.

FACTS 2.7.52.

(1) The field F is contained in an algebraically closed field E.
(2) There is an extension E of F that is both algebraically closed and algebraic over F.
(3) Suppose F is a field and E is an algebraic extension of F. Let σ be a monomorphism (injective homomorphism) of F into an algebraically closed field L. Then σ can be extended to a monomorphism of E into L.
(4) If L and L' are algebraically closed fields that are algebraic over F, then there exists an isomorphism $\tau : L \longrightarrow L'$ such that τ is the identity on F.

EXERCISE 2.7.53. Prove the above statements. Use Lang [**14**, Chapter 5] if you must.

EXERCISE 2.7.54. Show that if F is algebraically closed, then the algebraic closure of F is F.

REMARK 2.7.55. The Fundamental Theorem of Algebra shows that $\mathbb{C}$ is algebraically closed and, in fact, that $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$.

EXERCISE 2.7.56.

(i) Show that $\mathbb{A}_{\mathbb{R}}$, the field of real algebraic numbers, is not algebraically closed.

(ii) Show that $\mathbb{A}$, the field of algebraic numbers, is the algebraic closure of $\mathbb{Q}$.

2.7.3. Other Complete Ordered Fields.

Are there complete ordered fields for which the least upper bound property fails? Yes, and we will construct one.

The polynomial ring $\mathbb{R}[x]$ is not a field. However, we can construct a field from $\mathbb{R}[x]$ in the same way we constructed $\mathbb{Q}$ from $\mathbb{Z}$ in Appendix A. Before we do that, we wish to define an order on the polynomial ring $\mathbb{R}[x]$.

DEFINITION 2.7.57. Let $p(x)$ and $q(x)$ be elements of $\mathbb{R}[x]$. We say that $q(x) < p(x)$ if the leading coefficient of $p(x) - q(x)$ is a positive real number. Thus, a nonzero polynomial $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ in $\mathbb{R}[x]$ is positive if $a_n > 0$ in $\mathbb{R}$.

EXERCISE 2.7.58. Show that the order relation defined above makes $\mathbb{R}[x]$ into an ordered integral domain.

Define a nonempty subset A of $\mathbb{R}[x]$ to be *bounded above* if there exists $M \in \mathbb{R}[x]$ such that $a \leq M$ for all $a \in A$. Similarly, A is *bounded below* if there exists $L \in \mathbb{R}[x]$ such that $a \geq L$ for all $a \in A$. A *bounded* set is a set that is bounded above and below.

Life in $\mathbb{R}[x]$ is quite interesting. For example, $p(x) = x$ is positive and is greater than any real number.

EXERCISE 2.7.59. Prove that, with the above order, the set $\mathbb{R}$ has no least upper bound in $\mathbb{R}[x]$.

In $\mathbb{R}[x]$, it remains true that the really small elements are the really small real numbers. If we have a sequence $(p_k(x))_{k\in\mathbb{N}}$ in $\mathbb{R}[x]$, we say that $(p_k(x))_{k\in\mathbb{N}}$ is a *Cauchy sequence* if, given $\varepsilon \in \mathbb{R}$ such that $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that, for $n, m \geq N$, we have $-\varepsilon < p_n(x) - p_m(x) < \varepsilon$.

EXERCISE 2.7.60.

(i) Show that every Cauchy sequence in $\mathbb{R}[x]$ converges, so that $\mathbb{R}[x]$ is a complete ordered integral domain in which the least upper bound property does not hold.

(ii) If we define order on $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$ as above, we will again get ordered integral domains. Are these ordered integral domains complete?

So what? We are really looking for complete ordered fields in which the least upper bound property does not hold. A complete ordered integral domain in which the least upper bound property does not hold, as above, is a weak substitute. Here is where the fun begins.

Let R be a field, and consider $R[x]$, the ring of polynomials in one variable over R. Set

$$F = \{(p(x), q(x)) \mid p(x), q(x) \in R[x] \text{ and } q(x) \neq 0\}.$$

For $(p(x), q(x))$, $(r(x), s(x)) \in F$, we define $(p(x), q(x)) \sim (r(x), s(x))$ if $p(x)s(x) = r(x)q(x)$.

Exercise 2.7.61.

(i) Show that $\sim$ is an equivalence relation on F.
(ii) Show that, with addition and multiplication defined as usual, the set of equivalence classes forms a field exactly as we did in Appendix A when we constructed $\mathbb{Q}$ from $\mathbb{Z}$.

Definition 2.7.62. The field of equivalence classes constructed above is called the *quotient field* of $R[x]$, or *the field of rational functions* over R. This field is denoted by $R(x)$. The equivalence class of the pair $(p(x), q(x))$ is often written as the quotient $p(x)/q(x)$.

Exercise 2.7.63. Show that if R is an ordered field, then $R(x)$ is an ordered field with the definition $p(x)/q(x) > 0$ if $p(x)q(x) > 0$.

We now take $R = \mathbb{R}$ and make a few observations about $\mathbb{R}(x)$, the field of rational functions over $\mathbb{R}$ with the above order.

Facts 2.7.64.

(1) The field $\mathbb{R}$ is a subfield of $\mathbb{R}(x)$, and the order on $\mathbb{R}$ inherited from $\mathbb{R}(x)$ is the usual order on $\mathbb{R}$.
(2) If $r(x)$ and $s(x)$ are rational functions in $\mathbb{R}(x)$ and $r(x) < s(x)$, we define open intervals in the usual way, that is, $(r(x), s(x)) = \{u(x) \in \mathbb{R}(x) \mid r(x) < u(x) < s(x)\}$. We can also define closed and half-open intervals similarly.
(3) The small elements in $\mathbb{R}(x)$ are not the same as the small numbers in $\mathbb{R}$. For example, the rational function $1/x$ is positive and less than any positive real number. This last statement is clear since, if $r \in \mathbb{R}$ and $r > 0$, then $r < x$. Hence, $1/x < 1/r$.

Exercise 2.7.65. Given $p(x)/q(x) \in \mathbb{R}(x)$ with $p(x)/q(x) > 0$, show that there exists $N \in \mathbb{N}$ such that $1/x^N < p(x)/q(x)$.

Definition 2.7.66. We can define a topology on $\mathbb{R}(x)$ as follows. If $U \subseteq \mathbb{R}(x)$, then U is *open* if, given a point $r(x) \in U$, there is an $N \in \mathbb{N}$ such that the interval $(r(x) - 1/x^N, r(x) + 1/x^N) \subseteq U$. The collection of these sets is a topology on $\mathbb{R}(x)$ called the *order topology*.

EXERCISE 2.7.67. Suppose that $r \in \mathbb{R}$. Show that $(r-1/x, r+1/x) \cap \mathbb{R} = \{r\}$.

It follows from the last exercise that the relative topology (see the project in Section 2.7.1) on $\mathbb{R}$ inherited from $\mathbb{R}(x)$ is the discrete topology. This means that the copy of the real numbers contained in $\mathbb{R}(x)$ is not the usual real numbers that we have come to know and love. The usual topology on $\mathbb{R}$ generated by the ordinary open intervals certainly is not the discrete topology. For example, in the usual topology, the sequence $(1/n)_{n\in\mathbb{N}}$ is a Cauchy sequence that converges to 0. In the discrete topology inherited from $\mathbb{R}(x)$, this sequence is not a Cauchy sequence and so does not converge to anything at all. Even though $\mathbb{R}$ as a subset of $\mathbb{R}(x)$ is an ordered field and **in this field** the least upper bound property holds, this copy of $\mathbb{R}$ does not have the topological properties discussed in Chapter 1.

The topology on $\mathbb{R}(x)$ can also be generated by a metric.

DEFINITION 2.7.68. Let $p(x)/q(x) \in \mathbb{R}(x)$. We define an absolute value on $\mathbb{R}(x)$ as follows:

$$|p(x)/q(x)| = \begin{cases} 2^{\deg p - \deg q}, & p(x)/q(x) \neq 0, \\ 0, & p(x)/q(x) = 0. \end{cases}$$

EXERCISE 2.7.69. Show that $|\cdot|$ defined above satisfies:

(i) $|p(x)/q(x)| \geq 0$ and $|p(x)/q(x)| = 0$ if only if $p(x)/q(x) = 0$.
(ii) $\left|\frac{p(x)}{q(x)} \cdot \frac{r(x)}{s(x)}\right| = \left|\frac{p(x)}{q(x)}\right| \left|\frac{r(x)}{s(x)}\right|$.
(iii) $\left|\frac{p(x)}{q(x)} + \frac{r(x)}{s(x)}\right| \leq \max\left(\left|\frac{p(x)}{q(x)}\right|, \left|\frac{r(x)}{s(x)}\right|\right)$ with equality if $\left|\frac{p(x)}{q(x)}\right| \neq \left|\frac{r(x)}{s(x)}\right|$.

With this absolute value, we can define a metric on $\mathbb{R}(x)$ by setting

$$d\left(\frac{p(x)}{q(x)}, \frac{r(x)}{s(x)}\right) = \left|\frac{p(x)}{q(x)} - \frac{r(x)}{s(x)}\right|.$$

We will encounter this same kind of absolute value when we meet the p-adic numbers in Chapter 3.

EXERCISE 2.7.70. Show that the open sets in this metric topology are the same as the open sets in the order topology.

The fact is that $\mathbb{R}(x)$ is not complete with respect to this metric. That is, we can find Cauchy sequences in $\mathbb{R}(x)$ that do not converge. For example, consider the sequence

$$s_n = 1 + \frac{1}{x^{a_1}} + \cdots + \frac{1}{x^{a_n}}$$

in $\mathbb{R}(x)$ where $(a_k)_{k\in\mathbb{N}}$ is an increasing sequence of natural numbers. It is easy to see that this sequence is Cauchy. As we will see in the next proposition, it is possible to impose conditions on the sequence $(a_k)_{k\in\mathbb{N}}$ so that the sequence $(s_n)_{n\in\mathbb{N}}$ does not converge in $\mathbb{R}(x)$.

PROPOSITION 2.7.71. *Suppose that the sequence* $(s_n)_{n\in\mathbb{N}}$ *converges to the fraction* $p(x)/q(x) \in R(x)$ *and suppose that* $\deg q(x) = d$. *Then* $a_k \leq kd$ *for all* $k \geq 1$.

Proof. Suppose that $(s_n(x))_{n\in\mathbb{N}}$ converges to $p(x)/q(x)$ with $\deg q(x) = d$, and suppose on the contrary that there exists $k \geq 1$ such that $a_k > kd$. Let k_0 denote the smallest such k.

We know that, for all $N \in \mathbb{N}$, we have

$$-1/x^N < s_n - p(x)/q(x) < 1/x^N$$

for all sufficiently large n. Rewriting this, we get the pair of conditions

$$\frac{q(x)x^{a_n} - x^N[q(x)(1 + x^{a_n-a_{n-1}} + \cdots + x^{a_n}) - x^{a_n}p(x)]}{x^{a_n+N}q(x)} > 0,$$

$$\frac{-q(x)x^{a_n} - x^N[q(x)(1 + x^{a_n-a_{n-1}} + \cdots + x^{a_n}) - x^{a_n}p(x)]}{x^{a_n+N}q(x)} < 0.$$

Examining the degrees of the summands in the numerators, this immediately implies that

$$a_n + d \geq N + \deg[q(x)(1 + x^{a_n-a_{n-1}} + \cdots + x^{a_n}) - x^{a_n}p(x)]. \tag{2.1}$$

Now choose $n \geq k_0$ and let us rewrite the the polynomial whose degree appears above as

$$\begin{aligned} &q(x)(1 + x^{a_n-a_{n-1}} + \cdots + x^{a_n}) - x^{a_n}p(x) \\ &\quad = q(x)(1 + x^{a_n-a_{n-1}} + \cdots + x^{a_n-a_{k_0}}) \\ &\qquad + q(x)(x^{a_n-a_{k_0-1}} + \cdots + x^{a_n}) - x^{a_n}p(x). \end{aligned}$$

The largest degree occurring in the first summand after the equal sign is $d + a_n - a_{k_0}$, while the smallest possible degree occurring in the remaining two terms is $a_n - a_{k_0-1}$. But

$$d + a_n - a_{k_0} < d + a_n - dk_0 = a_n - (k_0 - 1)d \leq a_n - a_{k_0-1}$$

by minimality of k_0. It follows that the degree of the entire expression above is equal to the degree of the latter part

$$q(x)(x^{a_n-a_{k_0-1}} + \cdots + x^{a_n}) - x^{a_n}p(x)$$

since this polynomial cannot be zero (otherwise we would have $s_{k_0-1} = p(x)/q(x)$). As we have observed, the latter degree is at least $a_n - a_{k_0-1} \geq a_n - (k_0 - 1)d$. Returning to the inequality (2.1), we conclude that $a_n + d \geq N + a_n - (k_0 - 1)d$, which is to say $N \leq k_0 d$. But this is to hold for all $N \geq 1$, which is absurd.

REMARK 2.7.72. One consequence of the proposition above is that, by choosing an increasing sequence $(a_k)_{k\in\mathbb{N}}$ of natural numbers such that a_k/k is unbounded, we will have a Cauchy sequence $(s_n(x))_{n\in\mathbb{N}}$ in $\mathbb{R}(x)$ that does not converge in $\mathbb{R}(x)$.

EXERCISE 2.7.73.

(i) Show that $\sum_{k=0}^{\infty} \frac{1}{x^{k^2}}$ diverges in $\mathbb{R}(x)$.
(ii) Show that the least upper bound property does not hold in $\mathbb{R}(x)$.
(iii) Show that, in the completion of $\mathbb{R}(x)$, the least upper bound property does not hold, thus producing a complete ordered field in which the least upper bound property does not hold.
(iv) Find the algebraic closure of $\mathbb{R}(x)$.

Chapter 3

Complete Metric Spaces

> ...I have tried to draw the conclusions from the developments of the last thirty years, whereby locally compact groups, measure and integration have been seen to play an increasingly important role in classical number theory. ...In retrospect, we see now that the real numbers appear there as one of the infinitely many completions of the prime field, one which is neither more nor less interesting to the arithmetician than its p-adic companions, and that there is at least one language and one technique, that of the adèles, for bringing them all together under one roof and making them cooperate for a common purpose.
>
> – *André Weil*
> *Basic Number Theory*

Challenge Problems: III

PROBLEM 3.1. Suppose X_1 and X_2 are compact metric spaces. Show that $\mathcal{C}(X_1, \mathbb{R})$ is isometric to $\mathcal{C}(X_2, \mathbb{R})$ if and only if X_1 is homeomorphic to X_2.

PROBLEM 3.2. Suppose A is a subset of $\mathbb{R}^2$. Show that A can contain at most one point p such that A is isometric to $A \setminus \{p\}$ with the usual metric.

PROBLEM 3.3. Show that any continuous function $f : S^n \longrightarrow \mathbb{R}^n$ maps two antipodal points to the same point in $\mathbb{R}^n$, that is, there exists $x \in S^{n-1}$ so that $f(x) = f(-x)$.

PROBLEM 3.4. The positive integer n is a Borsuk number if any bounded set $S \subseteq \mathbb{R}^n$ can be partitioned into $n + 1$ subsets of smaller diameter. The integers 2 and 3 are Borsuk numbers. It is known that sufficiently large integers are not Borsuk numbers, and the upper bound is getting smaller annually. Find those $n \geq 4$ that are Borsuk numbers.

PROBLEM 3.5. Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ have the property that $\|f(x) - f(y)\| = \|x - y\|$ for all $x, y \in \mathbb{R}^n$. Assume that $f(0) = 0$. Show that f is linear.

In this chapter, we study some interesting ideas related to the concept of completeness in metric spaces. These ideas are all basic tools in analysis and play a significant role in the remainder of this text.

These notions include:

(1) the Contraction Mapping Theorem and its applications to differential equations;
(2) the Baire Category Theorem and its consequences;
(3) the Stone-Weierstrass theorem;
(4) the p-adic completion $\mathbb{Q}_p$ of $\mathbb{Q}$ for a given prime p.

3.1. The Contraction Mapping Theorem and Its Applications to Differential and Integral Equations

DEFINITION 3.1.1. Let X be a metric space and let f be a map from X to X. We say that f is a *contraction mapping of* X if there exists a real number α, with $0 < \alpha < 1$, such that $d(f(x), f(y)) \leq \alpha d(x, y)$ for every pair $x, y \in X$.

EXERCISE 3.1.2. Show that a contraction mapping is continuous.

DEFINITION 3.1.3. Let X be a metric space and let f be a map from X to X. A point $x_0 \in X$ is a *fixed point* of f if $f(x_0) = x_0$.

EXERCISE 3.1.4.

(i) Let $f : [0, 1] \longrightarrow [0, 1]$ be continuous. Show that f has a fixed point.
(ii) Find a continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ that does not have a fixed point.

THEOREM 3.1.5 **(Contraction Mapping Theorem)**. *Let X be a complete metric space and let $f : X \longrightarrow X$ be a contraction mapping. Then f has a unique fixed point $x_0 \in X$.*

Proof. Let x_1 be any element of X. Define $x_2 = f(x_1)$, $x_3 = f(x_2) = f(f(x_1)) = f^2(x_1)$, and in general $x_n = f^{n-1}(x_1)$. Then, if $n > m$, we have

$$\begin{aligned} d(x_m, x_n) &= d(f^{m-1}(x_1), f^{n-1}(x_1)) \\ &\leq \alpha^{m-1} d(x_1, f^{n-m}(x_1)) \\ &\leq \alpha^{m-1}(d(x_1, x_2) + d(x_2, x_3) + \cdots + d(x_{n-m}, x_{n-m+1})) \\ &\leq \alpha^{m-1}(d(x_1, x_2) + \alpha d(x_1, x_2) + \cdots + \alpha^{n-m-1} d(x_1, x_2)) \\ &\leq \frac{\alpha^{m-1}}{1-\alpha} d(x_1, x_2). \end{aligned}$$

It follows that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in X that converges since X is complete. Let $x_0 = \lim_{n\to\infty} x_n$. From the continuity of f, it follows that $f(x_0) = x_0$.

EXERCISE 3.1.6. Show that x_0 is the unique fixed point of f.

EXERCISE 3.1.7.

(i) Let $B = B_1(0)$ be the unit ball in the usual metric on $\mathbb{R}^n$, and let f be a map from B to B. We say that f satisfies a *Lipschitz condition* if there exists a constant C such that $|f(x) - f(y)| \leq C|x - y|$ for all $x, y \in B$. Show that if $0 < C < 1$, then f is a contraction mapping. Show that if $C \geq 1$, then f need not be a contraction mapping.

(ii) Let $T : \ell_n^p(\mathbb{R}) \longrightarrow \ell_n^q(\mathbb{R})$, $1 \leq p, q \leq \infty$, be a linear transformation. When is T a contraction mapping?

The following theorem, sometimes called *Picard's theorem*, gives a direct application of the contraction mapping theorem to a problem in analysis.

THEOREM 3.1.8 **(Picard's theorem)**. *Let B be a ball of radius r in $\mathbb{R}^2$ with center at (x_0, y_0). Suppose that $f : B \longrightarrow \mathbb{R}$ is a continuous function that satisfies a Lipschitz condition in the second variable; that is, there is a constant C such that $|f(x, y) - f(x, y')| \leq C|y - y'|$ for all $(x, y), (x, y') \in B$. Then there exists a $\delta > 0$ such that the differential equation $dy/dx = f(x, y)$ has a unique solution $y = \phi(x)$ in the interval $|x - x_0| < \delta$ satisfying $\phi(x_0) = y_0$.*

Proof. Without loss of generality, we can assume that f is bounded on B, that is, there exists a constant M so that $|f(x, y)| \leq M$ for all $(x, y) \in B$. Take a $\delta > 0$ such that $M\delta < 1$ and

$$\{(x, y) \mid |x - x_0| < \delta, |y - y_0| < M\delta\} \subseteq B.$$

We now work inside the space $X = \{\phi \in \mathcal{C}([x_0 - \delta, x_0 + \delta]) \mid |\phi(x) - y_0| \leq M\delta\}$. If we define a metric on X with the sup norm, then X is a complete metric space. Now, take the mapping $T : X \longrightarrow X$ defined by

$$T\phi(x) = y_0 + \int_{x_0}^{x} f(t, \phi(t))\, dt.$$

It is obvious that $T\phi \in X$ and that $\|T\phi - T\phi'\| < M\delta\|\phi - \phi'\|$, where $\|\cdot\|$ represents the sup norm. Thus, T is a contraction mapping on X, and there is a unique function $\phi \in X$ so that $T\phi = \phi$. It is easy to check that this ϕ is a solution to the original differential equation that satisfies the initial condition $\phi(x_0) = y_0$.

The Contraction Mapping Theorem can also be applied to systems of differential equations; see for example [**12**].

3.2. The Baire Category Theorem and the Uniform Boundedness Principle

The notion of Baire category is fundamental in several parts of analysis, in particular, in measure theory and the study of bounded linear operators on Banach spaces. The word category should not be confused with the

notions of "categories and functors". These notions occur in many parts of mathematics; however, they will not be found in this book.

We begin with some simple ideas in metric spaces.

DEFINITION 3.2.1. Let X be a metric space. A subset $A \subseteq X$ is said to be *nowhere dense* if $\complement(\bar{A})$ is dense in X. A set $A \subseteq X$ is said to be of the *first category* if A is a countable union of nowhere dense sets. A set $A \subseteq X$ is said to be of the *second category* if A is not of the first category.

EXAMPLES 3.2.2.

(i) Let $A = \mathbb{Z}^n$, the set of integer lattice points in $\mathbb{R}^n$ with the usual metric. Then A is nowhere dense and hence of the first category.
(ii) Let $X = [0,1] \subseteq \mathbb{R}$ with the usual metric, and let A be the Cantor set. Then A is nowhere dense in $[0,1]$.
(iii) In a nonempty discrete metric space, the only nowhere dense set is the empty set.
(iv) A set containing a single point in a metric space X is nowhere dense if and only if the point is not an isolated point of X.

EXERCISE 3.2.3.

(i) Show that a subset A of a metric space X is nowhere dense iff the interior of $\bar{A}$ is empty.
(ii) Suppose U is an open set in a metric space X. Show that $\bar{U} \setminus U$ is nowhere dense in X.
(iii) Suppose F is a closed subset of a metric space X. Show that $F \setminus F^\circ$ is nowhere dense in X.
(iv) Show that a countable union of sets of the first category is of the first category.

This is a good time to create a family of sets that is interesting for a number of purposes. The sets are called *generalized Cantor sets* and are a generalization of the Cantor set in Exercise 1.6.26. They are constructed as follows. Choose a real number α, with $0 < \alpha < 1$. Then delete successively from $[0,1]$ a sequence of open intervals. First, remove the open interval $(1/2 - \alpha/4, 1/2 + \alpha/4)$. From the two remaining closed intervals, remove the middle open intervals each of length $\alpha/8$. Continuing inductively, we remove a sequence of open intervals whose total length is $\sum_{n=1}^{\infty} \alpha/2^n = \alpha$.

The remaining set of points is denoted by C_α. If there is any justice in the world (and indeed there sometimes is), the "length" of C_α should be $1 - \alpha$. However, justice will not be served until Chapter 6.

DEFINITION 3.2.4. A subset A of a metric space is *perfect* if every point of A is an accumulation point of A.

REMARK 3.2.5. Note that we say in Exercise 1.6.26 that the ordinary Cantor set in the closed interval $[0,1]$ is a perfect set.

EXERCISE 3.2.6. For any α, with $0 < \alpha < 1$, show that C_α is a nowhere dense perfect set in $[0,1]$ with the usual metric.

In the idea of category, the crucial issue is not whether a given set is dense. For example, the rational numbers in $\mathbb{R}$ with the usual metric form a dense subset and are of the first category, while the unit interval in $\mathbb{R}$ is not dense in $\mathbb{R}$ and is of the second category.

EXERCISE 3.2.7.

(i) Show that the irrational numbers are of the second category in $\mathbb{R}$ with the usual metric.

(ii) Determine if each of the following sets is of the first or second category in $\mathbb{R}^n$ with the usual metric:
   (a) $A = \{(x_1, x_2, \ldots, x_n) \mid x_j \in \mathbb{Q} \text{ for } j = 1, 2, \ldots, n\}$,
   (b) $A = \{(x_1, x_2, \ldots, x_n) \mid \text{at least } k \text{ coordinates are rational for a fixed number } k\}$,
   (c) $A = \{(x_1, x_2, \ldots, x_n) \mid \text{at least 1 coordinate is irrational}\}$.

The following theorems of Baire are extremely important in analysis.

THEOREM 3.2.8 **(Baire)**. *Suppose that X is a complete metric space, and let $(U_n)_{n\in\mathbb{N}}$ be a collection of dense open subsets of X. Then $\bigcap_{n\in\mathbb{N}} U_n$ is dense in X.*

Proof. Let $U \subseteq X$ be a nonempty open set. We claim that $U \cap (\bigcap_{n\in\mathbb{N}} U_n) \neq \varnothing$. To show this, choose $x_1 \in U \cap U_1$ and $r_1 > 0$ so that $\bar{B}_{r_1}(x_1) \subseteq U_1 \cap U$. Next, since U_2 is open and dense, we can choose $x_2 \in B_{r_1}(x_1) \cap U_2 \cap U$ and r_2, with $0 < r_2 < r_1/2$ such that $\bar{B}_{r_2}(x_2) \subseteq B_{r_1}(x_1) \cap U_2$. Proceeding inductively, we choose $x_n \in B_{r_{n-1}}(x_{n-1})$ and r_n, with $0 < r_n < r_{n-1}/2$, such that $\bar{B}_{r_n}(x_n) \subseteq B_{r_{n-1}}(x_{n-1}) \cap U_n$. It is clear that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and, since X is complete, this sequence converges to a point x_0 in X. It follows immediately that x_0 is in U, and this proves the theorem.

COROLLARY 3.2.9 **(Baire Category Theorem)**. *Let X be a complete metric space. Then X is of the second category in itself.*

Proof. This is clear from the theorem above.

EXERCISE 3.2.10. Let X be a complete metric space and let $(F_n)_{n\in\mathbb{N}}$ be a countable collection of closed sets in X such that $X = \bigcup_{n\in\mathbb{N}} F_n$. Show that $U = \bigcup_{n\in\mathbb{N}} F_n^\circ$ is dense.

EXERCISE 3.2.11. Show that a complete metric space with no isolated points is uncountable.

A basic theorem which shows the usefulness of Baire category is the uniform boundedness principle. This is often thought of as a theorem about linear operators on Banach spaces, but it is really a theorem about continuous functions on complete metric spaces.

THEOREM 3.2.12 **(Uniform boundedness)**. *Let X be a complete metric space and let $S \subseteq \mathcal{C}(X, \mathbb{R})$. Suppose that, for each $x \in X$, there is a constant M_x such that $|f(x)| \leq M_x$ for all $f \in S$. Then there is a nonempty*

open subset $U \subseteq X$ and a constant M so that $|f(x)| \leq M$ for all $f \in S$ and all $x \in U$.

Proof. For $n \in \mathbb{N}$ and $f \in S$, let $S_{n,f} = \{x \in X \mid |f(x)| \leq n\}$. Let $S_n = \bigcap_{f \in S} S_{n,f}$. Since f is continuous, $S_{n,f}$ is closed, and hence, S_n is closed. For each $x \in X$, there is an n such that $|f(x)| < n$ for all $f \in S$. Hence $X = \bigcup_{n \in \mathbb{N}} S_n$. The completeness of X implies there exists an $N \in \mathbb{N}$ so that S_N is not nowhere dense. Since S_N is nonempty and closed, it contains an open ball U and, for each $x \in U$, $|f(x)| \leq N$ for all $f \in S$.

Exercise 3.2.13.

(i) Give an example of a set S of functions such that $\sup_{x \in X} M_x = \infty$.
(ii) In the previous problem, does the answer change if X is a compact metric space?

3.3. Stone-Weierstrass Theorem

Throughout this section F is $\mathbb{R}$ or $\mathbb{C}$.

Definition 3.3.1. Let A be a collection of functions from a set X to F. We say that the collection A *separates points* if, for every pair of distinct points $x_1, x_2 \in X$, there is a function $f \in A$ such that $f(x_1) \neq f(x_2)$.

Example 3.3.2. If $X = [0, 1]$, then $\mathcal{C}(X, \mathbb{R})$ separates points. This is easy to see just by drawing a picture.

Exercise 3.3.3.

(i) Show that polynomial functions in $\mathcal{C}([0, 1], \mathbb{R})$ separate points.
(ii) Can the polynomial functions in part (i) be taken with integer coefficients?

Definition 3.3.4. A real polynomial function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a finite linear combination of expressions of the form $x_1^{m_1} x_2^{m_2} \cdots x_n^{m_n}$ where $m_1, m_2, \ldots, m_n$ are nonnegative integers. The coefficients of a polynomial may be taken from $\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$. The resulting set of polynomials is denoted by $\mathbb{Z}[x_1, \ldots, x_n]$, $\mathbb{Q}[x_1, \ldots, x_n]$, and $\mathbb{R}[x_1, \ldots, x_n]$, respectively.

Example 3.3.5. A typical polynomial in $\mathbb{R}[x_1, x_2, x_3, x_4]$ looks like $\sqrt{2} x_1^3 x_2 x_3^2 x_4 + \pi x_1 x_2^5 x_4^{15} - 11 x_1 x_4$.

Exercise 3.3.6.

(i) Show that $R[x_1, x_2, \ldots, x_n]$ is a commutative ring with 1 for $R = \mathbb{Z}, \mathbb{Q}$, or $\mathbb{R}$. Find the units (invertible elements) in each of these rings.
(ii) Find the possible images of a polynomial in $\mathbb{R}[x_1, \ldots, x_n]$.

Theorem 3.3.7 **(Weierstrass)**. *Let A be a compact set in $\mathbb{R}^n$. Then every continuous function $f : A \longrightarrow \mathbb{R}$ is the uniform limit of a sequence of polynomials in $\mathbb{R}[x_1, \ldots, x_n]$.*

THEOREM 3.3.8 **(Stone)**. *Let X be a compact metric space. Let A be an algebra of continuous, real-valued functions on X, and suppose that A separates points. Then $\bar{A}$, the uniform closure of A, coincides either with $\mathcal{C}(X,\mathbb{R})$ or with $\mathcal{C}_{t_0} = \{f \in \mathcal{C}(X,\mathbb{R}) \mid f(t_0) = 0\}$, for some point $t_0 \in X$.*

Note that Stone's theorem implies Weierstrass's theorem simply by letting $A = \mathbb{R}[x_1, x_2, \ldots, x_n]$. Before we attempt the proof, it will be helpful to gather some preliminary lemmas.

LEMMA 3.3.9. *Let A be an algebra of real-valued, continuous functions on a compact metric space X. Then, for $f \in A$, $|f|$ is in the uniform closure of A. That is, there is a sequence $(f_n)_{n\in\mathbb{N}}$ in A so that $(f_n)_{n\in\mathbb{N}}$ converges uniformly to $|f|$.*

Proof. Since X is compact, we know that f is bounded. Choose $C \in \mathbb{R}$, $C > 0$, so that $|f(x)| \leq C$ for all $x \in X$. Let $u = \frac{1}{C}f$. Then $u \in A$ and $\|u\| \leq 1$. Now we construct a sequence $(w_n)_{n\in\mathbb{N}}$ in A converging uniformly to $|u|$.

Let $w_0 = 0$, and define w_n inductively by the relation

$$w_{n+1} = w_n + \frac{u^2 - w_n^2}{2}.$$

Before proceeding further, notice that if we formally take limits in n, we would have a relation of the form $w = w + (u^2 - w^2)/2$, which would imply that $w^2 = u^2$. With a little luck, we may also show that $w \geq 0$ and hence $w = |u|$.

First notice that $0 \leq w_1 - w_0 = w_1 = u^2/2 \leq u^2 \leq |u|$. Now suppose $w_k - w_{k-1} \geq 0$ and that $w_k \leq |u|$ for $1 \leq k \leq n$. Then,

$$w_{n+1} - w_n = \frac{u^2 - w_n^2}{2} = \frac{|u| + w_n}{2}(|u| - w_n) \geq 0.$$

Also, keeping in mind that $|u| \leq 1$, we have

$$\begin{aligned} 0 \leq w_{n+1} &= w_n + \frac{u^2 - w_n^2}{2} = w_n + \frac{|u| + w_n}{2}(|u| - w_n) \\ &\leq w_n + |u| - w_n = |u|. \end{aligned}$$

Hence, by induction, $(w_n)_{n\in\mathbb{N}}$ is an increasing sequence of functions and $0 \leq w_n \leq |u|$ for all n. Now, as suggested in the beginning of the proof, we let $w = \lim_{n\to\infty} w_n$ pointwise. Then, $w = |u|$, and by Dini's theorem, Exercise 2.6.11, we know that the sequence $(w_n)_{n\in\mathbb{N}}$ converges uniformly to $|u|$.

DEFINITION 3.3.10. Let V be a vector space of real-valued continuous functions on a metric space X. We say that V is a *lattice* if $|f| \in V$ whenever $f \in V$.

EXERCISE 3.3.11. Let V be a lattice on a metric space X. If f, g are in V, set $f \wedge g = \min(f, g)$ and $f \vee g = \max(f, g)$. Show that $f \wedge g$, $f \vee g \in V$.

LEMMA 3.3.12. *Let X be a compact metric space and let L be a lattice of continuous functions on X. Suppose that, for any $x, y \in X$ with $x \neq y$ and $a, b \in \mathbb{R}$, there is a function $f_{xy} \in L$ satisfying $f_{xy}(x) = a$ and $f_{xy}(y) = b$. Then, for each $f \in \mathcal{C}(X, \mathbb{R})$, there is a sequence $(f_n)_{n\in\mathbb{N}} \in L$ so that $(f_n)_{n\in\mathbb{N}}$ converges uniformly to f.*

Proof. Take $f \in \mathcal{C}(X, \mathbb{R})$ and $\varepsilon > 0$. For any $x, y \in X$, let $a = f(x)$ and $b = f(y)$ and find f_{xy} so that $f_{xy}(x) = a$ and $f_{xy}(y) = b$. We denote $U_{xy} = \{z \in X \mid f_{xy}(z) < f(z) + \varepsilon\}$ and $V_{xy} = \{z \in X \mid f(z) - \varepsilon < f_{xy}(z)\}$. Notice that for any $x, y \in X$, the sets U_{xy} and V_{xy} are open and, in addition, both contain x and y.

Fix y. Then by compactness, there exist a finite number of points $x_1, x_2, \ldots, x_n$ so that

$$\{U_{x_1y}, U_{x_2y}, \ldots, U_{x_ny}\}$$

covers X. Set $h_y = \min(f_{x_1y}, f_{x_2y}, \ldots, f_{x_ny})$. By Exercise 3.3.11, we have $h_y \in L$ and $h_y(z) < f(z) + \varepsilon$ for all $z \in X$. Now notice that $f(z) - \varepsilon < h_y(z)$ for $z \in V_y = \bigcap_{i=1}^{n} V_{x_iy}$.

Now let $y \in X$ vary, and for each y, construct h_y and V_y as above. By compactness, we can select an open cover $\{V_{y_1}, V_{y_2}, \ldots, V_{y_m}\}$ of X. Put $l = \max(h_{y_1}, h_{y_2}, \ldots, h_{y_m})$. Then $l \in L$ and $f(z) - \varepsilon < l(z) < f(z) + \varepsilon$.

Now, to construct $(f_n)_{n\in\mathbb{N}}$, we let $\varepsilon = 2^{-n}$ and choose f_n to be the function l constructed above.

We are ready to return to the proof of Stone's theorem.

Proof of Theorem 3.3.8. There are two cases to consider. First, suppose that, for each $x_0 \in X$, there is an $f \in A$ so that $f(x_0) \neq 0$. Take $x_1, x_2 \in X$ so that $x_1 \neq x_2$. Then there is a function $f \in A$ so that $f(x_1) \neq 0$ and $f(x_1) \neq f(x_2)$. To see this, take functions $h, g \in A$ so that $g(x_1) \neq g(x_2)$ and $h(x_1) \neq 0$. Then take

$$f(x) = \begin{cases} g(x) & \text{if } g(x_1) \neq 0, \\ h(x) & \text{if } g(x_1) = 0 \text{ and } h(x_1) \neq h(x_2), \\ g(x) + h(x) & \text{if } g(x_1) = 0 \text{ and } h(x_1) = h(x_2). \end{cases}$$

If $f(x_2) \neq 0$, let $u(x) = f(x)/f(x_2) - (f(x)/f(x_2))^2$. Then $u \in A$, $u(x_1) \neq 0$, and $u(x_2) = 0$.

Hence, we can find f_1 and f_2 in A so that $f_1(x_1) = 1$, $f_1(x_2) = 0$, $f_2(x_1) = 0$, and $f_2(x_2) = 1$. Now for any $a, b \in \mathbb{R}$ take $f = af_1(x) + bf_2(x)$. Then $f(x_1) = a$ and $f(x_2) = b$. From Lemma 3.3.9, we have that $\bar{A}$, the uniform closure of A, is a lattice. From Lemma 3.3.12, $\bar{A} = \mathcal{C}(X, \mathbb{R})$. This concludes the proof in the first case.

Now we turn to the case when there is an element $x_0 \in X$ so that $f(x_0) = 0$ for all $f \in A$. Let $A' = \{g \in \mathcal{C}(X, \mathbb{R}) \mid g(x) = c + f(x)$ for some $c \in \mathbb{R}$ and $f \in A\}$. We have that A' is an algebra satisfying the conditions for the first part of the theorem. In particular, if $h(x) \in \mathcal{C}_{x_0}(X, \mathbb{R})$ and $\varepsilon > 0$,

there is a function $f \in A$ and $c \in \mathbb{R}$ so that $\sup_{x\in X} |h(x) - c - f(x)| < \varepsilon$. Looking at x_0, we see that $|c| < \varepsilon$. Hence $\sup_{x\in X} |h(x) - f(x)| < 2\varepsilon$.

EXERCISE 3.3.13. Let X, Y be compact metric spaces. Show that the set $\{fg \mid f \in \mathcal{C}(X, \mathbb{R}) \text{ and } g \in \mathcal{C}(Y, \mathbb{R})\}$ is uniformly dense in $\mathcal{C}(X \times Y, \mathbb{R})$.

EXERCISE 3.3.14.

(i) Prove the complex version of the Stone-Weierstrass theorem. Let X be a compact metric space. Let A be an algebra of continuous complex-valued functions on X with the property that, if $f \in A$, then $\bar{f} \in A$. Assume that A separates points and that there is no point of $x \in X$ such that $f(x) = 0$ for all $f \in A$. Show that $\bar{A} = \mathcal{C}(X, \mathbb{C})$.

(ii) A *trigonometric polynomial* from $\mathbb{T}$ to $\mathbb{C}$ is a function of the form $f(\theta) = \sum_{-n}^{n} a_n e^{in\theta}$, where the coefficients are in $\mathbb{C}$. Show that the set of trigonometric polynomials is uniformly dense in $\mathcal{C}(\mathbb{T}, \mathbb{C})$.

3.4. The p-adic Completion of $\mathbb{Q}$

The simplest example of a completion of an incomplete metric space is called the p-adic completion of $\mathbb{Q}$. The p in this case refers to a prime integer, and the metric is that defined below. This metric plays a significant role in analysis, number theory, theoretical physics, and other areas.

The process of completing a metric space was carried out in Chapter 2. As mentioned there, the process for completing a metric space cannot be used to obtain the real numbers from the rational numbers because the method uses the fact that the real numbers are complete with respect to the usual metric.

3.4.1. Definitions and Basic Properties.

DEFINITION 3.4.1. Let p be a prime in $\mathbb{Z}$. For $r \in \mathbb{Q}^\times$, we write $r = p^k(\frac{a}{b})$ where a and b are relatively prime integers not divisible by p. Define the *p-adic absolute value* $|\cdot|_p$ on $\mathbb{Q}$ by

$$|r|_p = p^{-k} \quad \text{if } r \neq 0 \qquad \text{and} \qquad |0|_p = 0.$$

EXERCISE 3.4.2. Show that $|\cdot|_p$ has the following properties for all $r, s \in \mathbb{Q}$:

(i) $|r|_p \geqslant 0$, and $|r|_p = 0$ if and only if $r = 0$;
(ii) $|rs|_p = |r|_p \cdot |s|_p$;
(iii) $|r + s|_p \leq \max(|r|_p, |s|_p)$;
(iv) $|r + s|_p = \max(|r|_p, |s|_p)$ if $|r|_p \neq |s|_p$.

Note that (i) and (ii) are familiar properties of the usual absolute value on $\mathbb{Q}$, while (iii), known as *the non-Archimedean triangle inequality*, is stronger than the usual triangle inequality on $\mathbb{Q}$, that is,

$$|r + s| \leq |r| + |s|, \qquad r, s \in \mathbb{Q}.$$

The absolute value $|\cdot|_p$ gives a metric on $\mathbb{Q}$ defined by

$$d(r,s) = |r-s|_p, \qquad r,s \in \mathbb{Q}.$$

EXERCISE 3.4.3.

(i) Show that $|\cdot|_p$ defines a metric on $\mathbb{Q}$ as above.
(ii) Find a Cauchy sequence in $\mathbb{Q}$ relative to $|\cdot|_p$ that does not converge in $\mathbb{Q}$. Thus, $\mathbb{Q}$ is not complete with respect to $|\cdot|_p$.

We denote by $\mathbb{Q}_p$ the completion of $\mathbb{Q}$ with respect to the metric d. We can define addition and multiplication on $\mathbb{Q}_p$ so that $\mathbb{Q}_p$ becomes a field. Recall that elements of $\mathbb{Q}_p$ are equivalence classes of Cauchy sequences from $\mathbb{Q}$ relative to $|\cdot|_p$. The process of turning $\mathbb{Q}_p$ into a field proceeds exactly as in the case of the real numbers (see Section 1.5).

DEFINITION 3.4.4. Addition and multiplication on $\mathbb{Q}_p$ are defined as follows:

$$\begin{aligned} [a_n] + [b_n] &= [a_n + b_n], \\ [a_n] \cdot [b_n] &= [a_n b_n]. \end{aligned}$$

Next, we must extend $|\cdot|_p$ to $\mathbb{Q}_p$. Observe that, if $(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{Q}_p$, then $(|a_n|_p)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. So, if $[(a_n)_{n\in\mathbb{N}}] \in \mathbb{Q}_p$, then the absolute value on $\mathbb{Q}_p$ can be defined by

$$\big|[a_n]\big|_p = \lim_{n\to\infty} |a_n|_p.$$

Note that if $\lim_{n\to\infty} |a_n|_p \neq 0$, then the sequence $(|a_n|_p)_{n\in\mathbb{N}}$ is eventually constant and hence converges to the eventual constant.

EXERCISE 3.4.5.

(i) Show that addition, multiplication, and $|\cdot|_p$ are well-defined on $\mathbb{Q}_p$.
(ii) Show that $\mathbb{Q}_p$ is a field with the operations given above.
(iii) Show that $|\cdot|_p$ on $\mathbb{Q}_p$ satisfies the same properties as it does in $\mathbb{Q}$ (see Exercise 3.4.2).
(iv) Show that the image of $\mathbb{Q}_p$ under $|\cdot|_p$ is the same as that of $\mathbb{Q}$ under $|\cdot|_p$, that is, $\{p^k \mid k \in \mathbb{Z}\} \cup \{0\}$.
(v) Show that $\mathbb{Q}_p$ cannot be made into an ordered field.

DEFINITION 3.4.6. The field $\mathbb{Q}_p$ with $|\cdot|_p$ is called a *p-adic field*. It is also called the *p-adic completion of* $\mathbb{Q}$ or simply the *p-adic numbers*.

3.4.2. The Additive Structure of $\mathbb{Q}_p$.

We begin by defining several sets in $\mathbb{Q}_p$ that play an important role in our study of p-adic fields.

DEFINITION 3.4.7. Define the following subsets of $\mathbb{Q}_p$:

(a) $R_p = \{x \in \mathbb{Q}_p \mid |x|_p \leq 1\}$;
(b) $\mathfrak{p} = \{x \in R_p \mid |x|_p < 1\} = \{x \in R_p \mid |x|_p \leq \frac{1}{p}\}$; and
(c) $U_p = \{x \in R_p \mid |x|_p = 1\}$.

The set R_p is called the *ring of integers in* $\mathbb{Q}_p$. The set $\mathfrak{p}$ is called the *maximal ideal* in R_p. The set U_p is called the *group of units* in R_p.

There are several simple facts about the additive structure of $\mathbb{Q}_p$ related to the above subsets. We present some of them with proof; others we leave as exercises.

EXERCISE 3.4.8. Show that R_p is a commutative ring with 1.

PROPOSITION 3.4.9. *The set* $\mathfrak{p}$ *is a subgroup of* R_p*, and* $R_p = \bigcup_{0 \leq k \leq p-1} k + \mathfrak{p}$.

Proof. It follows from the non-Archimedean triangle inequality that $\mathfrak{p}$ is an additive subgroup of R_p. Let $x \in R_p$. If $|x|_p < 1$, then $x \in \mathfrak{p}$. Suppose $|x|_p = 1$. Since $\mathbb{Q}$ is dense in $\mathbb{Q}_p$, there is some $r \in \mathbb{Q}$ such that $r = a/b$ with $(a, b) = (a, p) = (b, p) = 1$ and $|r - x|_p < 1$. Hence, $x + \mathfrak{p} = r + \mathfrak{p}$. Since p and b are relatively prime, there exists an integer k with $0 < k \leq p - 1$ such that p divides $a - kb$. Hence, $|a - kb|_p < 1$, and also $|\frac{a-kb}{b}|_p < 1$ since $p \nmid b$. Thus, $|k - \frac{a}{b}|_p < 1$. It follows that $k + \mathfrak{p} = r + \mathfrak{p} = x + \mathfrak{p}$, so that $x \in k + \mathfrak{p}$.

EXERCISE 3.4.10.

(i) Show that U_p is, in fact, the set of *units* in R_p, that is, the set of elements in R_p that have multiplicative inverses in R_p.
(ii) Show that U_p is a group under multiplication.
(iii) Show that $\mathfrak{p}$ is an *ideal* in R_p, that is, if a is in $\mathfrak{p}$ and $x \in R_p$, then $ax \in \mathfrak{p}$.
(iv) Show that $\mathfrak{p}$ is the unique *maximal ideal* in R_p. See [**14**] for a discussion of ideals in commutative rings.
(v) For $n \in \mathbb{Z}$, define $\mathfrak{p}^n = p^n R_p = \{p^n x \mid x \in R_p\}$. Show that $\mathfrak{p}^n$ is a subgroup of $(\mathbb{Q}_p, +)$.
(vi) Show that $\mathfrak{p}^n \setminus \mathfrak{p}^{n+1} = p^n U_p$.
(vii) Show that if $n > 0$, $\mathfrak{p}^n$ is an ideal in R_p, that is, if $a \in \mathfrak{p}^n$ and $x \in R_p$, then $ax \in \mathfrak{p}^n$.
(viii) Show that $\mathbb{Q}_p = \bigcup_{n \in \mathbb{Z}} \mathfrak{p}^n$.
(ix) Show that $\mathbb{Q}_p^\times = \bigcup_{n \in \mathbb{Z}} p^n U_p$.

DEFINITION 3.4.11. If n is an integer, the set $p^n U_p$ is called a *shell* in $\mathbb{Q}_p$.

3.4.3. The Topological Structure of $\mathbb{Q}_p$.

We now consider the topology on $\mathbb{Q}_p$ determined by the metric associated to $|\cdot|_p$.

EXERCISE 3.4.12. If x_0 is an element of $\mathbb{Q}_p$ and $r > 0$, show that there is an integer n such that $B_r(x_0) = B_{p^{-n}}(x_0) = x_0 + \mathfrak{p}^{n+1} = \{x \in \mathbb{Q}_p \mid |x - x_0|_p < p^{-n}\} = \{x \in \mathbb{Q}_p \mid |x - x_0|_p \leq p^{-n-1}\}$.

This shows that the open balls in $\mathbb{Q}_p$ are simply cosets of some power of $\mathfrak{p}$.

PROPOSITION 3.4.13. *For each $n \in \mathbb{Z}$, the subsets $\mathfrak{p}^n$ and $p^n U_p$ are both open and closed in $\mathbb{Q}_p$.*

Proof. First, consider $p^n U_p$ for some $n \in \mathbb{Z}$. If $x \in p^n U_p$, then $|x|_p = p^{-n}$. If $k > n$, then the ball $x + \mathfrak{p}^k$ is contained in $p^n U_p$. This proves that $p^n U_p$ is open. Now, consider $\mathfrak{p}^n$. If $x \in \mathfrak{p}^n$ and $k > n$, then the ball $x + \mathfrak{p}^k$ is contained in $\mathfrak{p}^n$. Hence $\mathfrak{p}^n$ is open. To show that $\mathfrak{p}^n$ is closed, notice that $\mathbb{Q}_p \setminus \mathfrak{p}^n = \bigcup_{k<n} p^k U_p$, which is open. Finally, $p^n U_p$ is the complement of $\mathfrak{p}^{n+1}$ in $\mathfrak{p}^n$ so that $p^n U_p$ is closed.

COROLLARY 3.4.14. *If $n \in \mathbb{Z}$ and $x \in \mathbb{Q}_p$, then $x + \mathfrak{p}^n$ is both open and closed.*

COROLLARY 3.4.15. *Any open set A in $\mathbb{Q}_p$ can be written as a disjoint union of cosets of the subgroups $\mathfrak{p}^n$, $n \in \mathbb{Z}$.*

Proof. If A is empty, then we are done, so suppose that it is not. Suppose further that A is bounded. Then the set S of integers n such that A contains some coset of $\mathfrak{p}^n$ is bounded below. By the well-ordering principle we see that S has a least element, say n_0. Let $A_0 = a_0 + \mathfrak{p}^{n_0}$ be a coset of $\mathfrak{p}^{n_0}$ contained in A. By Corollary 3.4.14, A_0 is closed, so $A \setminus A_0$ is open. If $A \setminus A_0$ is empty, then we are done. Otherwise, repeat to get A_1, and so on.

EXERCISE 3.4.16. Prove that this algorithm terminates, so that we have written A as a disjoint union of finitely many cosets $A_0, A_1, \ldots, A_k$.

EXERCISE 3.4.17. Explain how to reduce the case of general A to the case of bounded A. (*Hint*: Consider the intersection of an arbitrary open set A with cosets of R_p.)

It is now easy to prove some of the basic topological properties of $\mathbb{Q}_p$. We present them in the form of exercises.

EXERCISE 3.4.18. Show that the ring of integers R_p is a maximal compact subring of $\mathbb{Q}_p$.

EXERCISE 3.4.19. Show that the field $\mathbb{Q}_p$ has the Bolzano-Weierstrass property, that is, if A is a bounded infinite subset of $\mathbb{Q}_p$, then A has an accumulation point in $\mathbb{Q}_p$.

EXERCISE 3.4.20. Show that $\mathbb{Q}_p$ has the Heine-Borel property, that is, if A is a closed, bounded subset of $\mathbb{Q}_p$, then A is compact.

EXERCISE 3.4.21. Show that the field $\mathbb{Q}_p$ is a locally compact field, that is, every point in $\mathbb{Q}_p$ has a neighborhood whose closure is compact. In fact, show that every point in $\mathbb{Q}_p$ has a neighborhood that is both open and compact.

The ring of p-adic integers modulo its maximal ideal, that is, $R_p/\mathfrak{p}$, is isomorphic to the finite field $\mathbb{F}_p$ with p elements. This is the same, of course, as $\mathbb{Z}_p$, the field of integers modulo p. This is a consequence of Proposition 3.4.9, which implies that

$$U_p = \bigcup_{1 \leqslant k \leqslant p-1} (k + \mathfrak{p}).$$

EXERCISE 3.4.22. Show that the set of cosets $k+\mathfrak{p}$ such that $0 \leq k \leq p-1$ can be made into a field which is isomorphic to $\mathbb{Z}_p$ by defining addition as usual and multiplication by $(k + \mathfrak{p})(j + \mathfrak{p}) = kj + \mathfrak{p}$.

EXERCISE 3.4.23. Show that $\mathbb{Z}_p^\times$, the multiplicative group of the field $\mathbb{Z}_p$, is a cyclic group.

Continuing this idea, we see that the shells of elements of $\mathbb{Q}_p$ with a particular absolute value may be written as

$$p^n U_p = \bigcup_{1 \leqslant k \leqslant p-1} (p^n k + \mathfrak{p}^{n+1}).$$

The next statement follows from the discussion just completed.

EXERCISE 3.4.24. The set of integers $\mathbb{Z}$ is dense in R_p relative to $|\cdot|_p$.

DEFINITION 3.4.25. The *valuation map* $\nu : \mathbb{Q}_p \longrightarrow \mathbb{Z} \cup \{+\infty\}$ is defined by the following rule:

$$p^{-\nu(x)} = |x|_p \quad \text{if } x \neq 0,$$
$$\nu(0) = +\infty.$$

With this definition, we can now write:

(a) $R_p = \{x \in \mathbb{Q}_p \mid \nu(x) \geq 0\}$;
(b) $\mathfrak{p} = \{x \in R_p \mid \nu(x) > 0\}$;
(c) $U_p = \{x \in R_p \mid \nu(x) = 0\}$.

We now consider the convergence of infinite series in $\mathbb{Q}_p$. The situation here is simpler than that in the project in Section 1.10.2.

In real and complex analysis, determining whether or not an infinite series converges can be a delicate matter. The p-adic case is different.

THEOREM 3.4.26. *Let $a_n \in \mathbb{Q}_p$ for all $n \in \mathbb{N}$. Then $\sum_{n=1}^{\infty} a_n$ converges in $\mathbb{Q}_p$ if and only if $\lim_{n\to\infty} a_n = 0$.*

Proof. The "only if" part is clear, just as in the real and complex cases. Now suppose that $\lim_{n\to\infty} a_n = 0$. This means that, given $k \geq 0$, we can pick $N \in \mathbb{N}$ such that $|a_n|_p < p^{-k}$ for all $n > N$. Thus, for all $m > n > N$,

$$|s_m - s_n|_p = |a_{n+1} + \cdots + a_m|_p \leq \max_{n+1\leq i\leq m} |a_i|_p < p^{-k},$$

the first inequality following from the non-Archimedean triangle inequality. Therefore, the sequence $(s_n)_{n\in\mathbb{N}}$ of partial sums is Cauchy and so must converge by the completeness of $\mathbb{Q}_p$.

From the decomposition $\mathbb{Q}_p^\times = \bigcup_{n\in\mathbb{Z}} p^n U_p$ of $\mathbb{Q}_p^\times$ into shells, one can express any nonzero x as an infinite series $x = \sum a_k p^k$, where the $a_k \in \{0, 1, \ldots, p-1\}$ are uniquely determined. In fact, the first nonzero term in the series is the one corresponding to the valuation of x, and one can write x in a *p-adic expansion*:

$$x = \sum_{k=\nu(x)}^{\infty} a_k p^k$$

where $a_{\nu(x)} \neq 0$. It follows immediately from Theorem 3.4.26 that the p-adic expansion of x converges.

EXERCISE 3.4.27. Show that the p-adic expansion of x converges to x.

DEFINITION 3.4.28. The *tail* of the p-adic expansion of x is defined as follows:

$$\lambda(x) = \begin{cases} \sum_{k=\nu(x)}^{-1} a_k p^k & \text{if } \nu(x) < 0, \\ 0 & \text{if } \nu(x) \geq 0. \end{cases}$$

So, $x \in R_p$ if and only if $\lambda(x) = 0$.

In Challenge Problem 4.4 we ask the reader to determine the properties of the p-adic expansions for rational numbers.

3.4.4. The Multiplicative Structure of $\mathbb{Q}_p$.

Given $x \in \mathbb{Q}_p^\times$, the valuation ν enables us to write $|x|_p = p^{-\nu(x)}$. Let $u = xp^{-\nu(x)}$. Then $|u|_p = 1$, and we can write $x = p^{\nu(x)}u$ for a unique $u \in U_p$. Hence $\mathbb{Q}_p^\times$ has a direct product decomposition as

$$\mathbb{Q}_p^\times \cong \{p^n \mid n \in \mathbb{Z}\} \times U_p \cong \mathbb{Z} \times U_p.$$

DEFINITION 3.4.29. If $n \in \mathbb{N}$, we define $U_{p,n} = 1 + \mathfrak{p}^n = \{x \in U_p \mid |1 - x|_p \leq p^{-n}\}$. For convenience, we write $U_{p,0} = U_p$.

Observe that $U_{p,1} = 1 + \mathfrak{p}$. This set is a group and is called the group of *principal units* in R_p. The $U_{p,n}$ are all subgroups of U_p under multiplication, and $U_p = U_{p,0} \supseteq U_{p,1} \supseteq U_{p,2} \supseteq \cdots \supseteq \{1\}$. The set $\{U_{p,n}\}_{n\in\mathbb{N}}$ is a neighborhood basis for the topology at 1.

It follows from the previous section that $U_p/U_{p,1} \cong \mathbb{F}_p^\times$, and the index of $U_{p,1}$ in U_p is $p-1$. More generally, the indices of these multiplicative subgroups are $[U_{p,n} : U_{p,m}] = p^{m-n}$ for $m \geqslant n \geqslant 1$ and $[U_p : U_{p,m}] = (p-1)p^{m-1}$. We now introduce a more refined direct product decomposition

$$\mathbb{Q}_p^\times \cong \mathbb{Z} \times U_p \cong \mathbb{Z} \times \mathbb{Z}_{p-1} \times U_{p,1}.$$

Recall that $\mathbb{F}_p^\times$ is a cyclic group. It turns out that we can lift a generator of $\mathbb{F}_p^\times$ to U_p.

PROPOSITION 3.4.30. *There exists $\varepsilon \in U_p$ such that $\varepsilon^{p-1} = 1$ and $\varepsilon^n \neq 1$ for $0 < n < p-1$. Moreover, U_p can be written as the disjoint union $U_p = U_{p,1} \cup \varepsilon U_{p,1} \cup \varepsilon^2 U_{p,1} \cup \cdots \cup \varepsilon^{p-2} U_{p,1}$.*

The standard way to prove this is to apply Hensel's lemma (see [**6**]). Another way is to provide a direct proof, which is essentially the proof of Hensel's lemma in this particular situation. We begin by considering the polynomial equation $f(X) = X^{p-1} - 1 = 0$. From the theory of finite fields, there exists an element α of $R_p/\mathfrak{p} \cong \mathbb{F}_p$ such that $\alpha^{p-1} = 1$ and such that $\alpha^n \neq 1$ for all $0 < n < p-1$. Our task then is to "lift" α to a suitable element in U_p.

We will instead apply a p-adic analog of Newton's method. This involves finding an approximate solution, showing that any approximate solution can be improved, and showing that the successive improvements converge to a solution.

First, we need an easy lemma.

LEMMA 3.4.31. *If $|1-x|_p = r < 1$, then $|1-x^{-1}|_p = r$.*

Proof. Exercise.

Proof of Proposition 3.4.30. Let $f(X)$ and α be as above. Choose any $a_0 \in U_{p,0} \setminus U_{p,1}$ such that $\eta(a_0) = \alpha$, where $\eta : R_p \to R_p/\mathfrak{p}$ is the natural map. Then $|f(a_0)|_p < 1$, so a_0 is an approximate zero for f.

Suppose $|f(b)|_p = r < 1$. Let $c = b - f(b)/f'(b)$, where f' is the derivative of f. We have

$$c = b - \frac{b^{p-1} - 1}{(p-1)b^{p-2}} = b\big(1 - \tfrac{1}{p-1}(1 - b^{1-p})\big),$$

so

$$\begin{aligned}
|f(c)|_p &= \left|b^{p-1}(1-\tfrac{1}{p-1}(1-b^{1-p}))^{p-1}-1\right|_p \\
&= \left|b^{p-1}\left(1-(1-b^{1-p})+\binom{p-1}{2}\left(\frac{1-b^{1-p}}{p-1}\right)^2+\cdots\right)-1\right|_p \\
&= \left|1+b^{p-1}\binom{p-1}{2}\left(\frac{1-b^{1-p}}{p-1}\right)^2+\cdots-1\right|_p \\
&= \left|\binom{p-1}{2}(1-b^{1-p})^2+\cdots\right|_p \\
&\leq r^2.
\end{aligned}$$

This last step follows from the non-Archimedean triangle inequality and the fact that every omitted term is a product of an integer and a power (of order at least two) of $(1-b^{1-p})$, which by the lemma has absolute value r. Thus, c is a better approximate solution than b, and

$$|c-b|_p = \left|\frac{b^{p-1}-1}{(p-1)b^{p-2}}\right|_p = |b^{p-1}-1|_p = r.$$

Now, use the procedure of the preceding paragraph to refine our approximate solution a_0 to a better solution a_1. Repeating this process, we get a Cauchy sequence $(a_n)_{n\in\mathbb{N}}$ in U_p. Let ε be the limit of this sequence in U_p. By the continuity of f, ε is a zero of f. Since $\eta(\varepsilon)=\alpha$, it is clear that $\varepsilon^n \neq 1$ for all $0<n<p-1$.

So ε is a primitive $(p-1)$-st root of unity in U_p. As stated in the proposition, the existence of such an ε allows us to write U_p as a disjoint union of multiplicative cosets

$$U_p = \bigcup_{k=0}^{p-2} \varepsilon^k U_{p,1}.$$

Note that $\varepsilon \notin U_{p,1}$. Otherwise, there would be some $x \in \mathfrak{p}\setminus\{0\}$ such that $(1+x)^{p-1}=1$, and this is not possible.

Exercise 3.4.32. Prove this last statement.

Returning to the additive structure for a moment, we can also write R_p as a disjoint union of additive cosets of the maximal ideal,

$$R_p = \mathfrak{p} \cup \bigcup_{k=0}^{p-2} (\varepsilon^k + \mathfrak{p}).$$

Next, we consider the structure of the group $(\mathbb{Q}_p^\times)^2$ of squares of elements of $\mathbb{Q}_p^\times$. In light of our direct product decomposition for $\mathbb{Q}_p^\times$, we must consider $U_{p,1}^2$.

PROPOSITION 3.4.33. *If p is odd, then $U_{p,1}^2 = U_{p,1}$.*

Proof. Let $\alpha \in U_{p,1}$. Then, as in the proof of Proposition 3.4.30, we are looking for a zero of a polynomial, in this case $f(X) = X^2 - \alpha$. As before, we start with an approximate solution and then use Newton's method to refine it to an exact one.

Define a sequence $(a_n)_{n\in\mathbb{N}}$ inductively by setting $a_0 = 1$ and

$$a_{i+1} = a_i - \frac{f(a_i)}{f'(a_i)}.$$

As in the proof of Proposition 3.4.30, this is a Cauchy sequence, and its limit is a square root of α.

EXERCISE 3.4.34. In the last proof, where did we use the fact that p was odd?

REMARK 3.4.35. If $p = 2$, then $U_{2,1}^2 = U_{2,3}$ (see [**28**]).

Any quadratic extension of $\mathbb{Q}_p$ has the form $\mathbb{Q}_p(\sqrt{\alpha})$ and is determined by the class of α in $\mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^2$, which must be nontrivial. If p is odd, then $(\mathbb{Q}_p^\times)^2 \cong 2\mathbb{Z} \times 2\mathbb{Z}_{p-1} \times U_{p,1}$. In particular, $[\mathbb{Q}_p^\times : (\mathbb{Q}_p^\times)^2] = 4$ in this case. Therefore, $\mathbb{Q}_p$ has three quadratic extensions, and they arise from adjoining $\sqrt{p}$, $\sqrt{\varepsilon}$, or $\sqrt{p\varepsilon}$ to $\mathbb{Q}_p$.

When $p = 2$, $U_{2,0} = U_{2,1}$, and we have the decomposition $\mathbb{Q}_2^\times \cong \mathbb{Z} \times U_{2,1}$. Hence, $(\mathbb{Q}_2^\times)^2 \cong 2\mathbb{Z} \times U_{2,3}$. Thus $[\mathbb{Q}_2^\times : (\mathbb{Q}_2^\times)^2] = 8$, and $\mathbb{Q}_2$ has seven distinct quadratic extensions.

3.5. Independent Projects

3.5.1. The Arzela-Ascoli Theorem.

Let F denote either $\mathbb{R}$ or $\mathbb{C}$.

DEFINITION 3.5.1. If X is a compact metric space and $S = \{f_i\}_{i\in I}$ is a collection of functions $f_i : X \longrightarrow F$, then S is *equicontinuous* if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $i \in I$, $d_X(x, y) < \delta$ implies that $|f(x) - f(y)| < \varepsilon$.

EXERCISE 3.5.2.

(i) If S in the above definition is a finite set of maps, show that S is equicontinuous if and only if all the f_i are continuous.
(ii) Construct an example (necessarily infinite) where all the f_i are continuous but S is not equicontinuous.

EXERCISE 3.5.3. If X is a compact metric space and $f_n \in \mathcal{C}(X)$ for $n \in \mathbb{N}$ and if $(f_n)_{n\in\mathbb{N}}$ converges uniformly on X, then $\{f_n\}_{n\in\mathbb{N}}$ is equicontinuous on X.

THEOREM 3.5.4 **(Arzela-Ascoli)**. *Suppose that $f_n \in \mathcal{C}(X)$ for $n \in \mathbb{N}$ and that $\{f_n\}_{n\in\mathbb{N}}$ is pointwise bounded and equicontinuous. Then $\{f_n\}_{n\in\mathbb{N}}$*

is uniformly bounded on X and $(f_n)_{n\in\mathbb{N}}$ contains a uniformly convergent subsequence.

Steps to the proof:

(1) Choose a countable dense subset $\{x_k \mid k \in \mathbb{N}\} \subseteq X$.
(2) Consider the sequence $f_n(x_1)$ and pick a convergent subsequence $f_{1,1}(x_1)$, $f_{1,2}(x_1), \ldots, f_{1,n}(x_1), \ldots$.
(3) Next pick a convergent subsequence $f_{2,1}(x_2), f_{2,2}(x_2), \ldots$ where the sequence $(f_{2,n})_{n\in\mathbb{N}}$ is a subsequence of $(f_{1,n})_{n\in\mathbb{N}}$.
(4) Proceeding in this fashion, we get a convergence sequence $f_{i,n}(x_i)$ such that $(f_{i+1,n})_{n\in\mathbb{N}}$ is a subsequence of $(f_{i,n})_{n\in\mathbb{N}}$ for all $i \geq 1$.
(5) Show that $f_{n,n}(x_k)$ converges as $n \to \infty$ for all $k \geq 1$.
(6) Complete the proof by showing that the subsequence $(f_{n,n})_{n\in\mathbb{N}}$ works.

3.5.2. Quadratic Extensions of $\mathbb{Q}_p$.

Let p be a prime in $\mathbb{Z}$. Recall that $\mathbb{Q}_p$ has either 3 or 7 distinct quadratic extensions, depending on whether p is odd or even, respectively. Suppose that $\theta \in \mathbb{Q}_p^\times \setminus (\mathbb{Q}_p^\times)^2$, so that

$$\mathbb{Q}_p(\sqrt{\theta}) = \{x + y\sqrt{\theta} \mid x, y \in \mathbb{Q}_p\}$$

is one such extension.

For $z = x + y\sqrt{\theta} \in \mathbb{Q}_p(\sqrt{\theta})$ we let $\overline{z} = x - y\sqrt{\theta}$ and refer to $\overline{z}$ as the *conjugate* of z.

Exercise 3.5.5. Show that the map $z \mapsto \overline{z}$ is an automorphism of the field $\mathbb{Q}_p(\sqrt{\theta})$ that leaves $\mathbb{Q}_p$ fixed.

We wish to define an absolute value on $\mathbb{Q}_p(\sqrt{\theta})$ that extends the usual absolute value $|\cdot|_p$ on $\mathbb{Q}_p$. The *relative field norm* for the extension $\mathbb{Q}_p(\sqrt{\theta})/\mathbb{Q}_p$ is defined by

$$\mathrm{Nm}_{\mathbb{Q}_p(\sqrt{\theta})/\mathbb{Q}_p}(z) = z\overline{z} = x^2 - \theta y^2$$

if $z = x + y\sqrt{\theta}$ as above.

Definition 3.5.6. The *absolute value* of $z \in \mathbb{Q}_p(\sqrt{\theta})$ is defined by

$$|z|_\theta = |\mathrm{Nm}_{\mathbb{Q}_p(\sqrt{\theta})/\mathbb{Q}_p}(z)|_p^{1/2} = |x^2 - \theta y^2|_p^{1/2}$$

if $z = x + y\sqrt{\theta}$.

Exercise 3.5.7. Show that $|\cdot|_\theta$ on $\mathbb{Q}_p(\sqrt{\theta})$ has the following properties:

(i) $|\cdot|_\theta$ extends the usual absolute value on $\mathbb{Q}_p$;
(ii) $|z|_\theta \geq 0$ and $|z|_\theta = 0$ iff $z = 0$;
(iii) $|zw|_\theta = |z|_\theta |w|_\theta$;
(iv) $|z + w|_\theta \leq \max(|z|_\theta, |w|_\theta)$;
(v) $|z + w|_\theta = \max(|z|_\theta, |w|_\theta)$ if $|z|_\theta \neq |w|_\theta$.

Thus $|\cdot|_\theta$ does indeed define a non-Archimedean absolute value on $\mathbb{Q}_p(\sqrt{\theta})$ that extends the usual one on $\mathbb{Q}_p$.

DEFINITION 3.5.8. The *group of norms* from $\mathbb{Q}_p(\sqrt{\theta})$ is defined by

$$\mathbb{Q}_{p,\theta}^{\times} = \mathrm{Nm}_{\mathbb{Q}_p(\sqrt{\theta})/\mathbb{Q}_p}(\mathbb{Q}_p(\sqrt{\theta})^{\times}).$$

EXERCISE 3.5.9.

(i) Show that $\mathbb{Q}_{p,\theta}^{\times}$ is a subgroup of $\mathbb{Q}_p^{\times}$ and that $[\mathbb{Q}_p^{\times} : \mathbb{Q}_{p,\theta}^{\times}] = 2$.
(ii) Show that $(\mathbb{Q}_p^{\times})^2$ is a subgroup of $\mathbb{Q}_{p,\theta}^{\times}$ and that $[\mathbb{Q}_{p,\theta}^{\times} : (\mathbb{Q}_p^{\times})^2]$ is 2 or 4 depending on whether p is odd or even, respectively.

DEFINITION 3.5.10. Let $\mathbb{Q}_p(\sqrt{\theta})$ be a quadratic extension of $\mathbb{Q}_p$. The *ring of integers* in $\mathbb{Q}_p(\sqrt{\theta})$ is

$$R_\theta = \{x \in \mathbb{Q}_p(\sqrt{\theta}) \mid |x|_\theta \leq 1\},$$

the *maximal ideal* in R_θ is

$$\mathfrak{p}_\theta = \{x \in \mathbb{Q}_p(\sqrt{\theta}) \mid |x|_\theta < 1\},$$

and the *unit group* of $\mathbb{Q}_p(\sqrt{\theta})$ is

$$U_\theta = \{x \in \mathbb{Q}_p(\sqrt{\theta}) \mid |x|_\theta = 1\}.$$

EXERCISE 3.5.11.

(i) Show that R_θ is a ring.
(ii) Show that $\mathfrak{p}_\theta$ is an ideal in R_θ.
(iii) Show that $\mathfrak{p}_\theta$ is the unique maximal ideal in R_θ.

From this point forward, we assume that p is odd. Thus, by the results in the chapter, we may take $\theta \in \{p, \varepsilon p, \varepsilon\}$.

EXERCISE 3.5.12.

(i) Show that $R_\theta = \{x + y\sqrt{\theta} \mid x, y \in R_p\}$.
(ii) If $\theta \in \{p, \varepsilon p\}$, show that $\mathfrak{p}_\theta = \{x + y\sqrt{\theta} \mid x \in \mathfrak{p}, y \in R_p\}$.
(iii) If $\theta = \varepsilon$, show that $\mathfrak{p}_\theta = \{x + y\sqrt{\theta} \mid x, y \in \mathfrak{p}\}$.

DEFINITION 3.5.13. We call the extensions $\mathbb{Q}_p(\sqrt{\theta})$ with $\theta \in \{p, \varepsilon p\}$ *ramified* and the extension $\mathbb{Q}_p(\sqrt{\varepsilon})$ *unramified*.

DEFINITION 3.5.14. The *residue field* of the extension $\mathbb{Q}_p(\sqrt{\theta})$ is the quotient $\mathbb{F}_\theta = R_\theta/\mathfrak{p}_\theta$.

EXERCISE 3.5.15. Show that $\mathbb{F}_\theta$ is an extension of $\mathbb{F}_p$ and that this extension has degree 1 or 2 depending on whether $\mathbb{Q}_p(\sqrt{\theta})$ is ramified or unramified, respectively.

3.5.3. Topological Groups.

3.5.3.1. Basic Properties

DEFINITION 3.5.16. A *topological group* is a triple $(G, \cdot, \mathcal{T})$ where

(a) $(G, \cdot)$ is a group,
(b) $(G, \mathcal{T})$ is a Hausdorff topological space, and
(c) the maps

$$\mu : G \times G \longrightarrow G \qquad i : G \longrightarrow G$$
$$(g, h) \longmapsto g \cdot h \qquad g \longmapsto g^{-1}$$

are continuous.

We will often omit the $\cdot$ and $\mathcal{T}$ from the notation and say "G is a topological group", as well as write the product of elements g and h simply as gh.

REMARK 3.5.17. Continuity of the group operation implies that, for all $g_0 \in G$, the map $G \longrightarrow G$ given by $g \mapsto g_0 g$ is a homeomorphism, since it has inverse given by $g \mapsto g_0^{-1} g$. These maps are called *left multiplication by g_0 and g_0^{-1}*, respectively. In particular, left multiplication by g_0 is an open and closed map. Similarly, we have homeomorphisms given by right multiplication.

This implies that G is a *homogeneous space*, that is, for all pairs of points $g, h \in G$, there exists a homeomorphism of G under which g is mapped to h. Indeed, left multiplication by hg^{-1} maps g to h.

Also note that inversion, i, is a homeomorphism and the inverse of i is in fact i.

From the definition of a topological group, one can immediately prove several properties about neighborhoods of points in such a group.

EXERCISE 3.5.18.

(i) Show that continuity of $\mu : G \times G \longrightarrow G$ implies that if $g, h \in G$, then for all neighborhoods W of gh, there exist neighborhoods U and V of g and h, respectively, such that $UV \subseteq W$.
(ii) Let $g \in G$. Show that, for every neighborhood V of g^{-1}, there exists a neighborhood U of g such that $U^{-1} \subseteq V$.
(iii) Let g, h be elements of G. Show that, for every neighborhood W of gh^{-1}, there exist neighborhoods U and V of g and h, respectively, such that $UV^{-1} \subseteq W$.
(iv) Let $g_1, \ldots, g_n \in G$, let $r_1, \ldots, r_n \in \mathbb{Z}$, and define c to be the product $g_1^{r_1} \cdots g_n^{r_n}$ of powers of the g_j's. Show that, for any neighborhood W of c, there exist neighborhoods $U_1, \ldots, U_n$ of $g_1, \ldots, g_n$, respectively, such that $U_1^{r_1} \cdots U_n^{r_n} \subseteq W$. Moreover, if $g_i = g_j$, we can arrange for U_i and U_j to be the same and similarly for larger numbers of equal elements.

PROPOSITION 3.5.19. *Suppose that $F \subseteq G$ is closed, $U \subseteq G$ is open, $h \in G$, and $S \subseteq G$ is any subset. Then hF, Fh, and F^{-1} are closed, and US, SU, and U^{-1} are open.*

Proof. The maps from G to G given by $g \mapsto hg$, $g \mapsto gh$, and $g \mapsto g^{-1}$ are homeomorphisms, hence closed, so hF, Fh, and F^{-1} are closed. Also, U^{-1} is open since the last map is a homeomorphism. Finally, $SU = \bigcup_{s \in S} sU$, and $US = \bigcup_{s \in S} Us$, which are open since $g \mapsto sg$ and $g \mapsto gs$ are open for all $s \in S$.

REMARK 3.5.20. Since a topological group G is homogeneous, it suffices to verify local properties at one point, usually the identity. As an example, we will show that all topological groups are regular.

EXERCISE 3.5.21. A Hausdorff topological space X is regular if and only if for all $x \in X$ and for all neighborhoods U of x there exists a neighborhood $V \subseteq U$ of x such that $\overline{V} \subseteq U$.

PROPOSITION 3.5.22. *Let G be a topological group. Then G is regular as a topological space.*

Proof. We will use the alternative characterization of regularity developed in the last exercise. Also, by the above remark, it suffices to consider neighborhoods of e, the identity. Let U be a neighborhood of e. Since $ee^{-1} = e$, by part (iv) of the Exercise 3.5.18, there exists a neighborhood V of e such that $VV^{-1} \subseteq U$. We will show that $\overline{V} \subseteq U$.

Let $p \in \overline{V}$. Then every neighborhood of p meets V. Since pV is open and since $p \in pV$ (remember that $e \in V$), there exists a point in $pV \cap V$, say $a \in V$, such that $a = pb \in pV$. Then, $p = ab^{-1} \in VV^{-1} \subseteq U$. So, $\overline{V} \subseteq U$.

EXERCISE 3.5.23. If $P, Q \subseteq G$ are compact, then so is PQ.

PROPOSITION 3.5.24. *Given any open set U containing e, there exists an open set $V \subseteq U$ such that $e \in V$ and $V = V^{-1}$.*

Proof. Take $V = U \cap U^{-1}$.

PROPOSITION 3.5.25. *If G is connected, then any neighborhood U of e generates G.*

Proof. By replacing U with $U \cap U^{-1}$ if necessary, we may assume that $U = U^{-1}$. Then, the group generated by U is $H := \bigcup_{n \in \mathbb{N}} U^n$, where $U^n = \{a_1 \cdots a_n \mid a_j \in U\}$. Since U is open, each U^n is open, and thus H is also open. Since H is open, it has open cosets, which partition G, so there must only be one coset and $H = G$.

3.5.3.2. Subgroups, Normal Subgroups, and Factor Groups

When we deal with subgroups of a topological group, we ordinarily require that they be closed in the topology of G. This will ensure that the coset space G/H is Hausdorff.

PROPOSITION 3.5.26. *If $H \subseteq G$ is a subgroup, then $\overline{H}$ is also a subgroup. Moreover, if H is normal, then so is $\overline{H}$. Finally, if H is an open subgroup in G, then H is also closed in G.*

Proof. Suppose that $a, b \in \overline{H}$. We want to show that $ab^{-1} \in \overline{H}$. Since $a, b \in \overline{H}$, every neighborhood of a intersects H and every neighborhood of b intersects H. Let W be a neighborhood of ab^{-1}. By Exercise 3.5.18, there exist neighborhoods U, V of a, b, respectively, such that $UV^{-1} \subseteq W$. Moreover, $U \cap H \neq \varnothing$ and $V \cap H \neq \varnothing$, so $V^{-1} \cap H \neq \varnothing$, which implies that $UV^{-1} \cap H \neq \varnothing$. Since W contains UV^{-1}, this means that $W \cap H \neq \varnothing$, and this implies that $ab^{-1} \in \overline{H}$.

Now, suppose that H is normal, and let $g \in G$, $h \in \overline{H}$. We want $ghg^{-1} \in \overline{H}$. Let W be a neighborhood of ghg^{-1}. By Exercise 3.5.18, there exist neighborhoods U of g and V of h such that $UVU^{-1} \subseteq W$. Now, since $h \in \overline{H}$, $V \cap H \neq \varnothing$ and we may take $h' \in V \cap H$. Since H is normal, $gh'g^{-1} \in H$ and moreover $gh'g^{-1} \in UVU^{-1}$. Thus $gh'g^{-1} \in H \cap UVU^{-1} \neq \varnothing$, and so $W \cap H \neq \varnothing$, and $ghg^{-1} \in \overline{H}$.

By Proposition 3.5.19, all of the cosets of an open subgroup H are open subsets of G. Thus, H is the complement of a union of open sets and therefore closed.

DEFINITION 3.5.27. Let G be a topological group and let $H \subseteq G$ be a closed subgroup. Then, the quotient space (with the quotient topology), G/H, is called the *space of left cosets of H in G*.

EXERCISE 3.5.28. G/H is a homogeneous space.

EXERCISE 3.5.29. If $q : X \longrightarrow Y$ is a quotient map, then $f : Y \longrightarrow Z$ is continuous if and only if $f \circ q : X \longrightarrow Z$ is continuous.

EXERCISE 3.5.30. If $N \triangleleft G$ is a closed, normal subgroup, then G/N is a topological group (with the quotient topology).

EXERCISE 3.5.31. If G is a compact topological group and $H \subseteq G$ is a closed subgroup, then H and G/H are compact.

3.5.3.3. Maps

DEFINITION 3.5.32. A *homomorphism* of topological groups G, G' is a continuous map $f : G \longrightarrow G'$ that is also a group homomorphism. An *isomorphism* of topological groups is a homomorphism with a continuous inverse.

PROPOSITION 3.5.33. *If $H \subseteq G$ is a closed subgroup, then the quotient map $q : G \longrightarrow G/H$ is open.*

Proof. Let $U \subseteq G$ be open. Then, $q(U) = \{uH \mid u \in U\}$. So, $q^{-1}(q(U)) = \{g \in G \mid gH = uH \text{ for some } u \in U\}$. Now, $gH = uH$ implies that $g = uh$ for some $h \in H$, so $q^{-1}(q(U)) = UH$, which is open, implying that $q(U)$ is open.

PROPOSITION 3.5.34. *If $f : G_1 \longrightarrow G_2$ is an open, surjective homomorphism, then $N = \ker(f)$ is a closed, normal subgroup and f induces an isomorphism $G_1/N \xrightarrow{\sim} G_2$.*

Proof. The subgroup N is normal since it is the kernel of a homomorphism. Since $N = f^{-1}(\{e\})$, N is closed. Let $\tilde{f} : G_1/N \longrightarrow G_2$ denote the induced map. Then, $\tilde{f}$ is continuous by Exercise 3.5.29.

Finally, that $\tilde{f}$ is open follows from the fact that f is open and q is continuous.

We conclude with an exercise and a useful proposition. This exercise takes a little bit of work.

EXERCISE 3.5.35. If $H \subseteq G$ is a closed subgroup, then G/H is regular.

Finally, we make an observation about locally compact groups, which are the ultimate domain of Fourier analysis.

PROPOSITION 3.5.36. *Suppose that G is a topological group and $H \subseteq G$ is a closed subgroup. Then if G is locally compact, so are H and G/H.*

Proof. The subgroup H is clearly locally compact. As usual, to show that G/H is locally compact, it suffices to show that $H \in G/H$ has a compact neighborhood. Let U be a compact neighborhood of the identity, e, in G. Then, $q(U) \subseteq G/H$ contains H and is compact, since q is continuous. Moreover, since q is open, $q(U)$ is a neighborhood of H, so G/H is locally compact.

Chapter 4

Normed Linear Spaces

> La théorie des opérations, créée par V. Volterra, a pour objet l'étude des fonctions définies dans les espaces à une infiniteé de dimensions. Dans plusieurs domaines très importants des mathématiques cette théorie a pénétré d'une facon essentielle: il suffit de rappeler que la théorie des équations intégrales et le calcul des variations se sont trouvés contenus comme des cas particuliers dans les principales sections de la théorie générale des opérations....
>
> Dans ce livre, je me suis proposé de recueillir surtout les résultats concernant les opérations linéaires définies dans certains espaces généraux, notamment dans les ainsi dits espaces du type (B), dont des cas particulier sont: l'espace des fonctions continues, celui des fonctions à p-ième puissance sommable, l'espace de Hilbert, etc.
>
> – *Stefan Banach*
> *Operation Linéaires*

Challenge Problems: IV

PROBLEM 4.1. Find $\mathrm{Aut}\,(\mathbb{Q}_p)$, the automorphism group of the field $\mathbb{Q}_p$.

PROBLEM 4.2. Describe the algebraic closure of $\mathbb{Q}_p$.

PROBLEM 4.3. Show that there cannot exist a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ that is continuous at the rational numbers and discontinuous at the irrational numbers.

PROBLEM 4.4. Show that a rational number has a periodic p-adic expansion, and determine the length of the period.

PROBLEM 4.5. An *absolute value* on $\mathbb{Q}$ is a function $|\cdot| : \mathbb{Q} \longrightarrow \mathbb{R}$ satisfying

(1) $|x| \geq 0$ and $|x| = 0$ iff $x = 0$,
(2) $|xy| = |x||y|$, and
(3) $|x + y| \leq |x| + |y|$.

Prove *Ostrowski's theorem*, which states that any absolute value on $\mathbb{Q}$ induces either the discrete topology, the usual topology, or the p-adic topology for some prime p.

4.1. Definitions and Basic Properties

As illustrated by Example 2.2.6, metric spaces become much more interesting when there is an algebraic structure accompanying the topological structure. In this chapter, we study vector spaces on which a norm is defined. This norm produces a metric with the usual properties. Throughout this chapter, the letter F denotes $\mathbb{R}$ or $\mathbb{C}$.

DEFINITION 4.1.1. Let V be a vector space over F. A *norm* on V is a map

$$\|\cdot\| : V \longrightarrow \mathbb{R}$$

such that:

(a) for all $v \in V$, $\|v\| \geq 0$, and $\|v\| = 0$ if and only if $v = 0$;
(b) for all $\alpha \in F$ and $v \in V$, $\|\alpha v\| = |\alpha|\|v\|$;
(c) for all v and v' in V, $\|v + v'\| \leq \|v\| + \|v'\|$ (the triangle inequality).

The pair $(V, ||\cdot||)$ is called a *real or complex normed linear space* when $F = \mathbb{R}$ or $\mathbb{C}$, respectively.

EXAMPLES 4.1.2.

(i) The spaces $\ell_n^p(\mathbb{R})$ are real normed linear spaces.
(ii) The spaces $\ell_n^p(\mathbb{C})$ are complex normed linear spaces.

For present purposes, we assume that $n \geq 2$. Note that, for $n = 1$, all these spaces are the same. The finite-dimensional spaces $\ell_n^p(F)$ have infinite-dimensional analogs.

DEFINITION 4.1.3. Suppose that $1 \leq p < \infty$. We set

$$\ell^p(F) = \left\{ x = (x_k)_{k\in\mathbb{N}} \;\middle|\; x_k \in F, \;\; \sum_{k=1}^{\infty} |x_k|^p < \infty \right\}.$$

The norm on $\ell^p(F)$ is defined by

$$||x||_p = \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{\frac{1}{p}}.$$

Here, the subscript p is not to be confused with the prime p in the p-adic metric. However, it should be noted that, among the choices of p with $1 \leq p < \infty$, all prime integers occur.

We also have an infinite-dimensional analog of the space $\ell_n^\infty(F)$.

DEFINITION 4.1.4. We define

$$\ell^\infty(F) = \left\{ x = (x_k)_{k\in\mathbb{N}} \,\middle|\, x_k \in F, \ \sup_{k\in\mathbb{N}} |x_k| < \infty \right\}.$$

The norm on $\ell^\infty(F)$ is defined by $\|x\|_\infty = \sup_k |x_k|$.

EXERCISE 4.1.5.

(i) Generalize Hölder's inequality to $\ell^p(F)$ for $1 \le p < \infty$.
(ii) Show that $\ell^p(F)$ is a normed linear space for $1 \le p \le \infty$.

REMARK 4.1.6. The reader may find this notation a bit quirky. When we dealt with finite-dimensional ℓ^p_n spaces, the dimension of the space was denoted by a subscript n. Now that we are working with infinite-dimensional spaces, we prefer not to use any subscript at all. Note that p can assume any value between 1 and ∞ inclusive. The most popular values of p are positive integers. However, we could also take $p = \pi$ or $p = e$, or other seemingly obscure values.

DEFINITION 4.1.7. Let V be a vector space over F with norm $\|\cdot\|$. We define a metric on V by setting $d(v, v') = ||v - v'||$. We call d *the metric induced by the norm.*

EXERCISE 4.1.8 **(Easy)**. Show that d as defined above is a metric on V.

DEFINITION 4.1.9. A normed linear space V is called a *Banach space* if V is complete with respect to the metric induced by the norm.

If you prefer, you may refer to a Banach space as a complete normed linear space.

EXAMPLES 4.1.10. Here are more examples of normed linear spaces. Recall the definitions of the spaces $\mathcal{B}(X, F)$ (Example 2.3.47) and $\mathcal{BC}(X, F)$ (Definition 2.4.31).

(i) If X is a set, $\mathcal{B}(X, F)$ is a normed linear space with the norm $\|f\| = \sup_{x\in X} |f(x)|$.
(ii) If X is a metric space, $\mathcal{BC}(X, F)$ is a normed linear space with the norm defined in (i). Note that if X is compact, we can simply write $\mathcal{C}(X, F)$.

DEFINITION 4.1.11. Let X be a metric space and let f be a function from X to F. We say that f is *compactly supported* if there is a compact set $B \subseteq X$ such that $f(x) = 0$ if $x \notin B$. In general, the *support* of a function $f : X \longrightarrow F$ is the closure of the set $\{x \in X \mid f(x) \neq 0\}$. The support of f is denoted by $\mathrm{supp} f$.

MORE EXAMPLES 4.1.12.

(iii) If X is a metric space, we set $\mathcal{C}_c(X, F) = \{f \in \mathcal{BC}(X, F) \mid f$ is compactly supported$\}$. This is a normed linear space with the norm defined in (i).

(iv) An important case of (iii) occurs when $X = \mathbb{N}$, with the discrete metric, and we get the space of all sequences that are zero except for a finite number of terms. This is the space $\mathcal{C}_c(\mathbb{N}, F)$. There are two additional sequence spaces.
(v) The first additional sequence space is $c_0(F) = \{(x_k)_{k\in\mathbb{N}} \in \ell^\infty(F) \mid \lim_{k\to\infty} x_k = 0\}$.
(vi) The second is $c(F) = \{(x_k)_{k\in\mathbb{N}} \in \ell^\infty(F) \mid (x_k)_{k\in\mathbb{N}} \text{ converges}\}$.

Note that the norm in the last three spaces is the ℓ^∞ norm.

EXERCISE 4.1.13. Show that all the examples above are normed linear spaces, and determine which among them are Banach spaces.

EXERCISE 4.1.14. Show that a finite-dimensional subspace of a normed linear space V is a closed subset of V.

An algebraic basis for any vector space is defined in Appendix B, even in the case when the space is infinite-dimensional. In the case of a normed linear space, we have a different notion of a basis that is related to the topology induced by the norm.

DEFINITION 4.1.15. Let $(V, \|\cdot\|)$ be a normed linear space. A *topological basis* for V is a linearly independent set whose finite linear span is dense in V.

EXAMPLE 4.1.16. For $j \in \mathbb{N}$, let the vector e_j be the infinite sequence of elements of F such that all the terms are zero, except for the j-th term, which is 1. The collection $\{e_j | j \in \mathbb{N}\}$ is a topological basis for $\ell^p(F)$, for any p with $1 \le p < \infty$.

EXERCISE 4.1.17. Prove this.

EXERCISE 4.1.18. Show that $\{e_j | \, j \in \mathbb{N}\}$ is not a topological basis for $\ell^\infty(F)$.

Let us consider topological bases for some of the examples above. The question of topological bases for Example 4.1.10(i) and (ii) and Example 4.1.12(iii) requires considerably more discussion. See for example [**15**], [**16**]. A particular case of Example 4.1.10(ii) is the case $X = [a, b] \subseteq \mathbb{R}$ and $F = \mathbb{C}$. In this case, the set of monomials $B = \{x^j \mid j = 0, 1, 2, \ldots\}$ forms a topological basis for $\mathcal{C}([a, b], \mathbb{C})$. This is the classical form of the Stone-Weierstrass theorem proved in Chapter 3.

The question that arises is, "What is the 'size' of an algebraic basis for a complete normed linear space?" The answer is contained in the following theorem.

THEOREM 4.1.19. *Let V be a Banach space. Then an algebraic basis for V is either finite or uncountable.*

Proof. Suppose V has a countable algebraic basis, $\{v_1, v_2, \ldots, v_n, \ldots\}$. Let V_n be the subspace of V spanned by $\{v_1, v_2, \ldots, v_n\}$. Then V_n is a closed

set in V, and $V = \bigcup_{n=1}^{\infty} V_n$. By the Baire Category Theorem, there exists an $N \in \mathbb{N}$ such that V_N has a nonempty interior. It then follows that $V_N = V$.

EXERCISE 4.1.20. Verify the last claim in the preceding proof.

REMARK 4.1.21. In the case of normed linear spaces, the most important subspaces are those that are closed in the topological sense. When we talk about a closed subspace of a normed linear space, we mean that it is closed in the sense of the topology, not just in the sense of the algebraic operations.

EXAMPLE 4.1.22.

(i) Any finite-dimensional subspace of a normed linear space is a closed subspace.
(ii) The set of all functions in $\mathcal{B}(X, F)$ that vanish on a fixed subset of X is a closed subspace.
(iii) The set of all elements $(a_n)_{n\in\mathbb{N}} \in \ell^p(F)$ such that $a_n = 0$ for n even is a closed subspace.

THEOREM 4.1.23. *Let V be a normed linear space and let V_0 be a closed subspace of V. If u is an element of V/V_0, that is, a coset in V/V_0, we define*

$$\|u\| = \operatorname{glb}\{\|v\| \mid v \in u\}.$$

Then $\|\cdot\|$ is a norm on the vector space V/V_0.

Proof. It is clear that $\|u\| \geq 0$ for all $u \in V/V_0$. If $\|u\| = 0$, then there exists a sequence $(v_n)_{n\in\mathbb{N}}, v_n \in u$, such that $\lim_{n\to\infty} \|v_n\| = 0$, which means that $\lim_{n\to\infty} v_n = 0$. Since V_0 is a closed subspace of V, it follows that every coset of V_0 is also closed (since translation by an element of V is an isometry). Hence $0 \in u$, and $u = 0$ in V/V_0. It is also obvious that $\|\alpha u\| = |\alpha|\|u\|$ for any $\alpha \in F$. Finally, to prove the triangle inequality, let $u, u' \in V/V_0$, and let $\varepsilon > 0$. Then there exist elements $v \in u$ and $v' \in u'$ such that $\|v\| \leq \|u\| + \varepsilon/2$ and $\|v'\| \leq \|u'\| + \varepsilon/2$. We have $v + v' \in u + u'$, whence

$$\|u + u'\| \leq \|v + v'\| \leq \|v\| + \|v'\| \leq \|u\| + \|u'\| + \varepsilon.$$

As this holds for every $\varepsilon > 0$, we get the triangle inequality.

THEOREM 4.1.24. *If V is complete, then V/V_0 is complete.*

Proof. Suppose that V is complete. Take a Cauchy sequence $(u_n)_{n\in\mathbb{N}}$ in V/V_0. Given $\varepsilon > 0$, we can take a subsequence $(u_k)_{k\in\mathbb{N}}$, such that $\|u_k - u_{k+1}\| < \varepsilon/2^{k+1}$ (see Exercise 1.6.13). We can choose $v_k \in u_k$ inductively for all k such that $\|v_k - v_{k+1}\| \leq \|u_k - u_{k+1}\| + \varepsilon/2^{k+1} \leq \varepsilon/2^k$. To see this, choose $w_1 \in u_1$ and $w_2 \in u_2$ such that $\|w_1 - w_2\| \leq \|u_1 - u_2\| + \varepsilon/4 \leq \varepsilon/2$. We may do this by the definition of the norm on V/V_0. Set $v_1 = w_1$ and $v_2 = w_2$. Continuing, we take $w_2' \in u_2$ and $w_3 \in u_3$ such that $\|w_2' - w_3\| \leq \varepsilon/4$ and we set $v_3 = w_3 + (v_2 - w_2')$. Notice that $v_3 \in u_3$ because $v_2 - w_2' \in V_0$. We continue inductively in this manner. It now follows that, for $j > k$, we have

$\|v_k - v_j\| \leq \varepsilon/2^{k-1}$ for $k \in \mathbb{N}$. So the sequence $(v_k)_{k\in\mathbb{N}}$ is Cauchy in V. Let $v_0 = \lim_{k\to\infty} v_k$.

EXERCISE 4.1.25. Show that

$$\lim_{n\to\infty} u_n = u_0,$$

where $u_0 = v_0 + V_0$.

EXERCISE 4.1.26.

(i) If V_0 is closed and V/V_0 is complete, does it follow that V is complete?
(ii) What happens if V_0 is not a closed subspace?
(iii) Let V and W be normed linear spaces and let $T : V \longrightarrow W$ be a linear map. Suppose T has the property that, for any ball B centered at the origin in V, the image $T(B)$ contains a ball centered at the origin in W. Show that T is an open map (see the project in Section 2.7.1).
(iv) If V is a normed linear space and V_0 is a closed subspace, show that the natural map from V to V/V_0 is an open map.

4.2. Bounded Linear Operators

The study of normed linear spaces is interesting, but the action really starts when we study bounded linear operators from one normed linear space to another. These operators will occur frequently throughout the remainder of the book and, before getting deeply involved in the analysis that lurks just over the horizon, we must develop their properties carefully.

DEFINITION 4.2.1. Let V and W be normed linear spaces over F. A linear map $T : V \longrightarrow W$ is a *bounded linear operator* if there exists a real number $M > 0$ such that $\|Tv\| \leq M\|v\|$ for all $v \in V$.

As in the definition, it is customary to write the action of a bounded linear operator on a particular vector without putting parentheses around the vector. We will follow this custom unless parentheses are required for clarity. We will have loads of examples of bounded linear operators, but, before that, we make a simple observation.

THEOREM 4.2.2. *Let V and W be normed linear spaces. Suppose that $T : V \longrightarrow W$ is a linear transformation. Then the following are equivalent:*

(i) *T is a bounded linear operator;*
(ii) *T is continuous;*
(iii) *T is continuous at one point.*

Proof. Suppose $\|Tv\| \leq M\|v\|$ for all $v \in V$. Then, given $\varepsilon > 0$ and $v, v' \in V$ such that $\|v - v'\| < \varepsilon/M$, we have that $\|Tv - Tv'\| < \varepsilon$. This proves (ii). Of course, (ii) implies (iii). Now assume T is continuous at v_0. Put $\varepsilon = 1$. Then there is a constant $\delta > 0$ such that

$$||Tv - Tv_0|| = ||T(v - v_0)|| < \varepsilon = 1$$

when $||v - v_0|| < \delta$. In turn, this implies that if $||v'|| < \delta$, then $||Tv'|| < 1$. Now suppose $v \neq 0$. Then,

$$||Tv|| = \frac{||v||}{\delta} \left\| T\left(\frac{\delta v}{||v||}\right) \right\| \leq \frac{||v||}{\delta}.$$

Take $M = \frac{1}{\delta}$.

EXAMPLE 4.2.3.

(i) Let $V = \ell^p(F)$. Fix an element $(b_n)_{n\in\mathbb{N}} \in \ell^\infty(F)$. Define $T : V \longrightarrow V$ by $T(a_n)_{n\in\mathbb{N}} = (a_n b_n)_{n\in\mathbb{N}}$. Here we can take $M = ||(b_n)_{n\in\mathbb{N}}||_\infty$.

(ii) Let $V = \ell^1(F)$. Fix an element $(b_n)_{n\in\mathbb{N}} \in \ell^\infty(F)$. Define $T : V \longrightarrow F$ by

$$T(a_n)_{n\in\mathbb{N}} = \sum_{n=1}^{\infty} a_n b_n.$$

Here again, we can take $M = ||(b_n)_{n\in\mathbb{N}}||_\infty$.

(iii) Let X be a metric space and fix $x_0 \in X$. Let $V = \mathcal{BC}(X, F)$ with the sup norm. Define $T_{x_0} : V \longrightarrow F$ by $T_{x_0}(f) = f(x_0)$. What can you take for M in this case?

Returning to the general theory of bounded linear operators (or transformations or maps, whichever you prefer), we wish to make a normed linear space out of the collection of all bounded linear operators from a normed linear space V to a normed linear space W.

DEFINITION 4.2.4. Let V and W be normed linear spaces over F. Let

$$\mathcal{BL}(V, W) = \{T : V \longrightarrow W \,|\, T \text{ is bounded and linear}\}.$$

For $T \in \mathcal{BL}(V, W)$, we define

$$\|T\| = \inf\{M \,|\, \|Tv\| \leq M\|v\| \text{ for all } v \in V\}.$$

EXERCISE 4.2.5.

(i) Show that $\mathcal{BL}(V, W)$ is a vector space over F.

(ii) Show that $||\cdot||$ defines a norm on $\mathcal{BL}(V, W)$.

(iii) Show that, for $V \neq 0$ and $T \in \mathcal{BL}(V, W)$,

$$||T|| = \sup_{v\neq 0} \frac{||Tv||}{||v||} = \sup_{||v||=1} ||Tv||.$$

THEOREM 4.2.6. *The space $(\mathcal{BL}(V, W), ||\cdot||)$ is a normed linear space. This space is complete if W is complete.*

Proof. The exercise above shows that $||\cdot||$ is a norm on $\mathcal{BL}(V, W)$. Suppose that W is complete and that $(T_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{BL}(V, W)$. Then, for every $v \in V$, $(T_n v)_{n\in\mathbb{N}}$ is a Cauchy sequence in W. Set $Tv = \lim_{n\to\infty} T_n v$. It is trivial to check that T is linear. We now prove that $T_n \to T$ uniformly (that is, $\|T_n - T\| \to 0$) and that T is bounded. Take $\varepsilon > 0$ and $N \in \mathbb{N}$ such that if $n, m > N$, then we have $\|T_n v - T_m v\| < \varepsilon$ whenever

$\|v\| = 1$. Taking the limit as $m \to \infty$, we deduce that $\|T_n v - Tv\| \leq \varepsilon$ for all $n \geq N$ and all v such that $\|v\| = 1$. The rest is easy.

If V is a Banach space, we denote by $\mathcal{BL}(V)$ the space $\mathcal{BL}(V, V)$. It turns out that, with composition of operators as multiplication, $\mathcal{BL}(V)$ is an algebra.

EXERCISE 4.2.7. If T_1 and T_2 are in $\mathcal{BL}(V)$, show that $\|T_1 T_2\| \leq \|T_1\| \, \|T_2\|$.

DEFINITION 4.2.8. Let X be an algebra that is also a normed linear space. The space X is called a *normed algebra* if the norm $\|\cdot\|$ satisfies

$$\|x_1 x_2\| \leq \|x_1\| \, \|x_2\| \quad \text{for all } x_1, x_2 \in X.$$

If X is complete with respect to the norm, then X is called a *Banach algebra*.

EXAMPLES 4.2.9.

(i) If V is a Banach space, then $\mathcal{BL}(V)$ is a Banach algebra.
(ii) If X is any metric space, then $\mathcal{BC}(X, F)$, with the sup norm, is a Banach algebra if multiplication is taken to be pointwise multiplication of functions.

EXERCISE 4.2.10.

(i) Show that any normed algebra can be isometrically embedded in another normed algebra that possesses a multiplicative identity.
(ii) Suppose V is a Banach space that is also an algebra such that the product $(x, y) \mapsto x \cdot y$ is continuous as a function of x and y separately. Show that there is an equivalent norm on V for which $\|x \cdot y\| \leq \|x\| \|y\|$.

Like any linear transformation, an element $T \in \mathcal{BL}(V, W)$ has a kernel, that is, $\ker T = \{v \in V \,|\, Tv = 0\}$. It is clear that $\ker T$ is a closed subspace of V. Thus, we have a natural linear map π from V to the normed linear space $V/\ker T$, where the norm on $V/\ker T$ is defined as in Theorem 4.1.23. In turn, we get the induced map $\tilde{T} : V/\ker T \longrightarrow W$.

EXERCISE 4.2.11. Show that $\|\tilde{T}\| = \|T\|$.

A situation of great interest is the case where $W = F$. Here, we consider continuous linear maps from V into F.

DEFINITION 4.2.12. The normed linear space $V^* = \mathcal{BL}(V, F)$ is called the *dual space* of V. The elements of V^* are often called *bounded linear functionals* on V.

Note that V^* is a Banach space since F is complete.

EXAMPLES 4.2.13.

(i) Let $V = \ell^1(F)$. Then $V^* = \ell^\infty(F)$. This is actually easy to show. Let $f \in \ell^1(F)^*$ and consider the action of f on the topological basis $(e_j)_{j\in\mathbb{N}}$. Let $b_j = f(e_j)$. Since f is bounded, there exists a constant M such that $|b_j| \leq M$ for all j. Thus, the sequence $(b_j)_{j\in\mathbb{N}}$ is in $\ell^\infty(F)$. It follows from the continuity of f that the action of f on $\ell^1(F)$ is given by $f((a_n)_{n\in\mathbb{N}}) = \sum_{n=1}^\infty a_n b_n$. On the other hand, if $(c_n)_{n\in\mathbb{N}}$ is an element of $\ell^\infty(F)$, then the map $f : \ell^1(F) \longrightarrow F$ defined by $f((a_n)_{n\in\mathbb{N}}) = \sum_{n=1}^\infty a_n c_n$ is an element of $\ell^1(F)^*$. To complete the identification of $\ell^1(F)^*$ with $\ell^\infty(F)$, we must show that the norm of f and the ℓ^∞ norm of $(c_k)_{k\in\mathbb{N}}$ coincide (see Exercise 4.2.16).

(ii) In Example 4.2.3, the map T_{x_0} is in $\mathcal{BC}(X, F)^*$. This map is called *pointwise evaluation.*

(iii) If $a = (a_n)_{n\in\mathbb{N}} \in \ell^1(F)$, then a can be regarded as an element in $\ell^\infty(F)^*$. This is defined by setting $f_a((b_n)_{n\in\mathbb{N}}) = \sum_{n=1}^\infty a_n b_n$. Note that in this case, the norm of f_a is less than or equal to $\|a\|_1$ since

$$\sum_{n=1}^{\infty} |a_n b_n| \leq \left(\sum_{n=1}^{\infty} |a_n| \right) \sup_{n\in\mathbb{N}} |b_n|.$$

EXERCISE 4.2.14. Show that, moreover, the norm of f_a is equal to $\|a\|_1$.

REMARK 4.2.15. In (ii) and (iii) above, the dual spaces contain much more than the examples indicate in general.

EXERCISE 4.2.16. In (i) above, show that $\|f\|$ (as an element of V^*) is equal to $\|(b_n)_{n\in\mathbb{N}}\|_\infty$.

This exercise illustrates an essential component in identifying the dual of a normed linear space. The process has two parts: first of all, identifying the elements in the dual space with elements of some normed linear space, and secondly, showing that the norm of a linear functional in V^* is the same as the norm of the element with which it is identified. One might say this identifies the abstract dual with the concrete dual.

EXERCISE 4.2.17.

(i) Let $V = \ell_n^p(F)$ where $1 < p < \infty$. Define q in $\mathbb{R}$ by the equation $1/p + 1/q = 1$. Show that $V^* = \ell_n^q(F)$.

(ii) Let $V = \ell^p(F)$ where $1 < p < \infty$. Define q in $\mathbb{R}$ by the equation $1/p + 1/q = 1$. Show that $V^* = \ell^q(F)$.

4.3. Fundamental Theorems about Linear Operators

In Chapter 3, based on the Baire Category Theorem, we proved the uniform boundedness principle. This is one of the most basic theorems in the theory of complete metric spaces, and it has extensive applications. Some theorems

that play fundamental roles in functional analysis follow from this principle. We present one of these now.

THEOREM 4.3.1 **(Banach-Steinhaus theorem)**. *Let V be a Banach space and let W be a normed linear space. Let $\{T_\alpha\}_{\alpha\in A}$ be a family of bounded linear operators from V to W. Suppose that, for each $v \in V$, there is a constant M_v such that $\|T_\alpha v\| \leq M_v$ for all $\alpha \in A$. Then there exists a constant M such that $\|T_\alpha\| \leq M$ for all $\alpha \in A$.*

Proof. For each $\alpha \in A$, define a function from V to $\mathbb{R}$ by $f_\alpha(v) = \|T_\alpha v\|$. Observe that each of the functions f_α is continuous. Thus, by Theorem 3.2.12, there is an open set $U \subseteq V$ and a constant M such that $|f_\alpha(v)| \leq M$ for all $\alpha \in A$ and $v \in U$. We want to extend this result from U to all of V. Let $\overline{B}_\delta(v_0)$ be a closed ball of radius δ centered at some element v_0 and contained in U. We know that $\|T_\alpha\| = \sup_{\|v\|=1} \|T_\alpha v\|$. So, for any v with $\|v\| = 1$, we write

$$T_\alpha v = \frac{1}{\delta}(T_\alpha(\delta v + v_0) - T_\alpha v_0).$$

Notice that $\delta v + v_0$ and v_0 are elements of $\overline{B}_\delta(v_0)$. Hence, $\|T_\alpha v\| \leq \frac{2M}{\delta}$. The right-hand side is independent of v. Thus, $\|T_\alpha\| \leq \frac{2M}{\delta}$.

This theorem is often called the *Uniform Boundedness Theorem*. The following three theorems are equivalent. We prove some of the equivalences and leave as exercises the remaining ones. We adopt the approach in Loomis [**17**].

THEOREM 4.3.2 **(Open Mapping Theorem)**. *Let V and W be Banach spaces and let $T \in \mathcal{BL}(V, W)$. If T is surjective, then T is an open map.*

THEOREM 4.3.3 **(Bounded Inverse Theorem)**. *Suppose that V and W are Banach spaces and $T \in \mathcal{BL}(V, W)$. If T is a bijection, then T^{-1} is in $\mathcal{BL}(W, V)$.*

THEOREM 4.3.4 **(Closed Graph Theorem)**. *Let V and W be Banach spaces and let T be a linear transformation from V to W. If the graph of T is closed in $V \times W$, then T is bounded.*

The fact that these three statements are equivalent is contained in the following propositions and exercises.

PROPOSITION 4.3.5. *The Bounded Inverse Theorem implies the Open Mapping Theorem.*

Proof. Suppose that V and W are Banach spaces and $T : V \longrightarrow W$ is a surjective bounded linear map. Let $V_0 = \ker T$. Then V_0 is a closed subspace of V, and the natural map π from V to V/V_0 is an open map (see Exercise 4.1.26). The induced map $\tilde{T} : V/V_0 \longrightarrow W$ is a bijective bounded linear transformation. By the Bounded Inverse Theorem, $\tilde{T}$ is an open mapping, and hence $T = \tilde{T} \circ \pi$ is an open mapping.

EXERCISE 4.3.6. Show that the Open Mapping Theorem implies the Bounded Inverse Theorem.

PROPOSITION 4.3.7. *The Bounded Inverse Theorem implies the Closed Graph Theorem.*

Proof. Suppose that V and W are Banach spaces and $T : V \longrightarrow W$ is a linear map such that the graph of T is closed in $V \times W$. The graph of T is a Banach space with the norm $\|(v, Tv)\| = \|v\| + \|Tv\|$. Since the map $(v, Tv) \mapsto v$ is a norm-decreasing bijective linear map onto V, it follows from the Bounded Inverse Theorem that $v \mapsto (v, Tv)$ is bounded, which implies that T is bounded.

EXERCISE 4.3.8. Prove that the Closed Graph Theorem implies the Bounded Inverse Theorem.

Now, we proceed to prove the Bounded Inverse Theorem following Loomis. We need two lemmas.

LEMMA 4.3.9. *Suppose that V and W are Banach spaces and $T : V \longrightarrow W$ is a bounded linear map. If the image under T of the unit ball $B = B_1(0)$ in V is dense in some ball $B_r = B_r(0)$ about the origin in W, then the image TB includes $B_r(0)$.*

Proof. We show that, for any ε with $0 < \varepsilon < 1$, $B_r \subseteq TB_{1/(1-\varepsilon)}(0)$. It then follows from linearity that for any ε with $0 < \varepsilon < 1$ the set $B_{r(1-\varepsilon)} \subseteq TB$ and hence $B_r \subseteq TB$. The set $A = B_r \cap TB$ is dense in B_r by assumption. Let w be any point in B_r. Given any ε, $0 < \varepsilon < 1$, we construct v in $B_{1/(1-\varepsilon)}(0) \subseteq V$ such that $Tv = w$. Take $w_0 = 0$ and inductively choose $w_n \in W$ so that $w_{n+1} - w_n \in \varepsilon^n A \subseteq A$ and $\|w_{n+1} - w\| < \varepsilon^n r$ for each $n \geq 0$. Then there exists a sequence $(v_n)_{n\in\mathbb{N}}$ such that $T(v_{n+1}) = w_{n+1} - w_n$ and $\|v_{n+1}\| \leq \varepsilon^n$. If we let $v = \sum_{n=1}^{\infty} v_n$, then $\|v\| < 1/(1-\varepsilon)$ and $T(v) = \sum_{n=1}^{\infty}(w_n - w_{n-1}) = w$; that is, the image of the set $B_{1/(1-\varepsilon)}(0)$ covers B_r. Thus $B_{r(1-\varepsilon)} \subseteq TB$ for every ε; hence $B_r \subseteq TB$.

LEMMA 4.3.10. *Suppose that V and W are Banach spaces and $T : V \longrightarrow W$ is a bounded linear map. If the image under T of the unit ball $B = B_1(0)$ in V is dense in no ball in W, then the range of T contains no ball in W.*

Proof. If TB is dense in no ball in W, then $TB_n = \{Tv \mid \|v\| < n\} = nTB$ has the same property. Given any ball $B_r \subseteq W$, there exists therefore a closed ball $\overline{B}_{r_1}(v_1) \subseteq B_r$ that is disjoint from TB, and then, by induction, a sequence of closed balls $\overline{B}_{r_n}(v_n) \subseteq \overline{B}_{r_{n-1}}(v_{n-1})$ such that $\overline{B}_{r_n}(v_n)$ is disjoint from TB_n. We can also require $r_n \to 0$, and then $(v_n)_{n\in\mathbb{N}}$ is Cauchy. The limit v lies in all the balls $\overline{B}_{r_n}(v_n)$ and hence in B_r, but in none of the sets TB_n. Since $\bigcup_{n=1}^{\infty} TB_n = TV$, we have proved that TV does not contain any ball $B_r \subseteq W$.

Proof (Bounded Inverse Theorem). Lemma 4.3.10 implies that $TB_1(0)$ is dense in some ball in W and therefore, by translation, that $TB_2(0)$ is dense in a ball $B_r(0)$. But then $B_r \subseteq TB_2(0)$ by Lemma 4.3.9, so $T^{-1}B_r \subseteq B_2(0)$, and $\|T^{-1}\| \leq 2/r$.

The following exercise runs along the same lines as we have used above. Nonetheless, the reader might find it useful to carry out the details.

EXERCISE 4.3.11. Let V be a Banach space, let W be a normed linear space, and let $T : V \longrightarrow W$ be a continuous linear mapping. Suppose that, for some bounded subset B of V, $T(B)$ is dense in the unit ball of W. Prove that $T(V) = W$, that T is an open mapping, and that W is complete.

4.4. Extending Linear Functionals

Let V be a finite-dimensional vector space over a field F, and let W be a proper subspace of V. It is shown in Appendix B that if f is a linear functional on W, that is, $f \in W^*$, then f can be extended to an element $F \in V^*$ with the property that $F|_W = f$. The extension F should not be confused with the field F, even though the letters look exactly the same. In the context of normed linear spaces, we would like to have a similar theorem. In this case, there should also be a statement related to the norms of the linear functionals. This is the content of the Hahn-Banach theorem.

THEOREM 4.4.1 **(Big Hahn-Banach)**. *Let V be a normed linear space, and let W be a subspace of V. If f is a bounded linear functional on W, then f can be extended to a bounded linear functional F on V such that $F|_W = f$ and $\|f\| = \|F\|$.*

We first prove a lemma that, in the case $F = \mathbb{R}$, allows us to extend linear functionals on W by one dimension.

LEMMA 4.4.2 **(Little Hahn-Banach)**. *Let V be a normed linear space over $\mathbb{R}$ and let W be a proper subspace of V. If f is a bounded linear functional on W and if $v_0 \in V \setminus W$, then there exists a bounded linear functional F on the subspace $W + \mathbb{R} \cdot v_0$, such that $F|_W = f$ and $\|f\| = \|F\|$.*

Proof. If $W = \{0\}$ or $f = 0$, then the result is trivial. Thus we may assume $\|f\| = 1$. For any elements $w_1, w_2 \in W$, we have

$$f(w_2) - f(w_1) = f(w_2 - w_1) \leq \|w_2 - w_1\| \leq \|w_2 - v_0\| + \|w_1 - v_0\|.$$

So, $-f(w_1) - \|w_1 - v_0\| \leq \|w_2 - v_0\| - f(w_2)$. It follows that

$$\sup_{w \in W} \left[-f(w) - \|w - v_0\|\right] \leq \inf_{w \in W} \left[\|w - v_0\| - f(w)\right].$$

We can choose c such that

$$\sup_{w \in W} \left[-f(w) - \|w - v_0\|\right] \leq c \leq \inf_{w \in W} \left[\|w - v_0\| - f(w)\right].$$

Set $F(v_0) = c$. Now, for $w \in W$ and $\lambda \in \mathbb{R}$, define $F(w + \lambda v_0) = f(w) + \lambda c$. It is clear that F is linear on $W + \mathbb{R} \cdot v_0$. Moreover, for any $w \in W$, we have $|F(w + \lambda v_0)| = |f(w) + \lambda c|$. From this, we get

$$|F(w + \lambda v_0)| = |\lambda||F(\lambda^{-1}w + v_0)| = |\lambda||f(\lambda^{-1}w) + c|.$$

By the inequality above, this is less than or equal to $|\lambda|\|\lambda^{-1}w + v_0\| = \|w + \lambda v_0\|$. Since it is clear that the norm of an extension is always greater than or equal to the norm of the original functional, we get $\|F\| = 1$.

Proof of Big Hahn-Banach over $\mathbb{R}$. Consider the collection of extensions of f, that is, pairs (W', F') such that $W \subseteq W' \subseteq V$, $F'|_W = f$, and $\|F'\| = \|f\|$. This collection is partially ordered by inclusion in the obvious sense, and, if we take a linearly ordered subfamily, then the union is an extension that is an upper bound for the subfamily. Here we take the linear functional on this union to be the one whose restriction to any member of the family is the linear functional associated with that member. By Zorn's lemma, there exists a maximal extension. By Little Hahn-Banach, this maximal extension must have domain V.

It is not difficult to extend the Big Hahn-Banach for real normed linear spaces to complex normed linear spaces.

EXERCISE 4.4.3. Let V be a normed linear space over $\mathbb{C}$, and let W be a subspace of V. If f is a bounded linear functional on W, show that there is a bounded linear functional F on V such that $F|_W = f$ and $\|F\| = \|f\|$. *Hint*: Extend the real part of f to a real linear functional G on V and set $F(v) = G(v) - iG(iv)$.

There is another version of the Hahn-Banach theorem that is often presented as the Hahn-Banach theorem. It is still about extending functionals but in a slightly different setting.

THEOREM 4.4.4 **(Other Hahn-Banach Theorem)**. *Let V be a vector space over $\mathbb{R}$, and let $p : V \longrightarrow \mathbb{R}$ be a function such that*

$$(**) \qquad p(x + y) \leq p(x) + p(y), \quad \text{for all } x, y \in V \text{ and}$$
$$p(\alpha x) = \alpha p(x), \quad \alpha \in \mathbb{R}, \alpha \geq 0.$$

Let W be a subspace of V, and let $f : W \longrightarrow \mathbb{R}$ be a linear functional that satisfies $f(x) \leq p(x)$ for all $x \in W$. Then there exists a linear functional $F : V \longrightarrow \mathbb{R}$ such that $F(x) = f(x)$ if $x \in W$ and $F(x) \leq p(x)$ for all $x \in V$. The function p is called a subadditive functional *from $V \longrightarrow \mathbb{R}$.*

Proof. Let Z be any subspace of V such that $W \subseteq Z \subseteq V$. Define

$$\mathcal{S} = \{(h, Z) \mid h \in \mathcal{L}(Z, \mathbb{R}), h(x) \leq p(x) \text{ for all } x \in Z \text{ and } h(x) = f(x) \text{ for all } x \in W\}.$$

Then $\mathcal{S}$ can be partially ordered by saying $(h, Z) \leq (h', Z')$ if $Z \subseteq Z'$ and h' extends h. By Zorn's lemma, $\mathcal{S}$ has a maximal element $(g, \tilde{Z})$. We claim

that $\tilde{Z}$ is all of V. If not, choose $y \in V \setminus \tilde{Z}$. Let $Y = \{x + ty | x \in \tilde{Z} \text{ and } t \in \mathbb{R}\}$. Clearly, Y is a subspace of V and we wish to show that g extends to $\tilde{g} : Y \longrightarrow \mathbb{R}$. This means we must find $c \in \mathbb{R}$ such that if we define $\tilde{g}$ by

$$\tilde{g}(x + ty) = g(x) + ct,$$

then

$$g(x) + ct \leq p(x + ty).$$

If $t > 0$, this means we need c to satisfy $ct \leq p(x + ty) - g(x)$, which is equivalent to requiring that

$$(\sharp) \qquad c \leq p\left(\frac{x}{t} + y\right) - g\left(\frac{x}{t}\right).$$

If $t < 0$, let $s = -t$. Then, showing that $ct \leq p(x+ty)-g(x)$ is equivalent to showing that $c(-s) \leq p(x - sy) - g(x)$. If we divide by $-s$, the requirement becomes

$$(\sharp\sharp) \qquad c \geq -p\left(\frac{x}{s} - y\right) + g\left(\frac{x}{s}\right).$$

To find c that satisfies ($\sharp$) and ($\sharp\sharp$), let $x, x' \in \tilde{Z}$. Then

$$g(x)+g(x') = g(x+x') \leq p(x+x') = p(x-y+y+x') \leq p(x-y)+p(x'+y).$$

If we rearrange this, we get

$$g(x) - p(x - y) \leq p(x' + y) - g(x').$$

The left-hand side, as a function of x, is bounded by the right-hand side (which is independent of x). Let c be the least upper bound of the left-hand side. This c satisfies ($\sharp$) and ($\sharp\sharp$).

EXERCISE 4.4.5. Suppose V is a Banach space over $\mathbb{R}$ and that p is a subadditive functional on V. Take $v_0 \neq 0$ in V and let $W = \{\alpha v_0 \mid \alpha \in \mathbb{R}\}$. Define a function f from W to $\mathbb{R}$ by $f(\alpha v_0) = \alpha p(v_0)$ for all $\alpha \in \mathbb{R}$. Show that $f(w) \leq p(w)$ for all $w \in W$. Hence by the Other Hahn-Banach Theorem, f extends to a bounded linear functional F on V such that $F(v) \leq p(v)$ for all $v \in V$.

EXERCISE 4.4.6. Show that Other Hahn-Banach implies Big Hahn-Banach.

4.5. Generalized Limits and the Dual of $\ell^\infty(F)$

In Example 4.2.13(i), we determined that the dual of $\ell^1(F)$ is $\ell^\infty(F)$. In Exercise 4.2.17, we indicated that for $1 < p < \infty$, the dual space of $\ell^p(F)$ is $\ell^q(F)$, where $1/p + 1/q = 1$. This leaves us with the problem of finding the dual of $\ell^\infty(F)$. In Example 4.2.13(iii), we showed that $\ell^1(F)$ is contained in the dual of $\ell^\infty(F)$. For the present, we will work with $\ell^\infty(\mathbb{R})$.

Note that if $(c_n)_{n\in\mathbb{N}} \in \ell^\infty(\mathbb{R})$, then $\limsup_{n\to\infty} |c_n|$ is finite, and in fact, $\limsup_{n\to\infty} |c_n| \leq \|(c_n)_{n\in\mathbb{N}}\|_\infty$. We next define a bounded linear functional on $\ell^\infty(\mathbb{R})$ that associates to any bounded sequence of real numbers $(c_n)_{n\in\mathbb{N}}$ a "generalized limit" denoted $\text{L. i. m.}_n\, c_n$.

This bounded linear functional has the following properties:

(1) If $c_n \geq 0$ for all $n \in \mathbb{N}$, then $\mathrm{L.\,i.\,m.}_n\, c_n \geq 0$.
(2) $\mathrm{L.\,i.\,m.}_n\, c_{n+1} = \mathrm{L.\,i.\,m.}_n\, c_n$.
(3) If $c_n = 1$ for all n, then $\mathrm{L.\,i.\,m.}_n\, c_n = 1$.

The process goes as follows. We first construct a subadditive functional on $\ell^\infty(\mathbb{R})$. We then proceed as in Exercise 4.4.5 using the Other Hahn-Banach Theorem to get a bounded linear functional on $\ell^\infty(\mathbb{R})$ that is dominated by our subadditive functional. To complete the construction, we must prove properties (1), (2), and (3) above.

We take an element $c = (c_n)_{n\in\mathbb{N}} \in \ell^\infty(\mathbb{R})$. To each finite sequence $i_1, \ldots, i_k$ of natural numbers, we assign the number

$$\limsup_{n\to\infty} \frac{1}{k} \sum_{j=1}^{k} c_{n+i_j}.$$

We define $p(c)$ to be the infimum of these numbers over all choices of finite sequences. The next step is to prove that p is subadditive.

PROPOSITION 4.5.1. *The functional p defined above is a subadditive functional on $\ell^\infty(\mathbb{R})$.*

Proof. First of all, it is obvious that $p(\alpha c) = \alpha p(c)$ for $\alpha \geq 0$. Let $c = (c_n)_{n\in\mathbb{N}}$ and $d = (d_n)_{n\in\mathbb{N}}$ be elements of $\ell^\infty(\mathbb{R})$. Let $\varepsilon > 0$ and choose finite sequences $i_1, \ldots, i_k$ and $j_1, \ldots, j_h$ such that

$$\limsup_{n\to\infty} \frac{1}{k} \sum_{m=1}^{k} c_{n+i_m} \leq p(c) + \varepsilon/2$$

and

$$\limsup_{n\to\infty} \frac{1}{h} \sum_{l=1}^{h} d_{n+j_l} \leq p(d) + \varepsilon/2.$$

We now turn to $p(c+d)$. We choose the finite sequence whose entries are of the form $i_m + j_l$ for $1 \leq m \leq k$ and $1 \leq l \leq h$. Then

$$\begin{aligned} p(c+d) &\leq \limsup_{n\to\infty} \frac{1}{kh} \sum_{m=1}^{k} \sum_{l=1}^{h} c_{n+i_m+j_l} + d_{n+i_m+j_l} \\ &= \limsup_{n\to\infty} \frac{1}{h} \sum_{l=1}^{h} \frac{1}{k} \sum_{m=1}^{k} c_{n+i_m+j_l} + \limsup_{n\to\infty} \frac{1}{k} \sum_{m=1}^{k} \frac{1}{h} \sum_{l=1}^{h} d_{n+i_m+j_l}. \end{aligned}$$

Now,

$$\limsup_{n\to\infty} \frac{1}{h} \sum_{l=1}^{h} \frac{1}{k} \sum_{m=1}^{k} c_{n+i_m+j_l} \leq \limsup_{n\to\infty} \frac{1}{k} \sum_{m=1}^{k} c_{n+i_m} \leq p(c) + \varepsilon/2.$$

A similar argument holds for the second term. Hence, $p(c+d) \leq p(c) + p(d) + \varepsilon$. Since ε was arbitrary, p is a subadditive functional on $\ell^\infty(\mathbb{R})$.

Now that we have a subadditive functional p on $\ell^\infty(\mathbb{R})$, we can construct a linear functional f dominated by p on $\ell^\infty(\mathbb{R})$ by Exercise 4.4.5.

To see that property (1) above holds for f, suppose $c_n \geq 0$ for all $n \in \mathbb{N}$. Then clearly $p(-c) \leq 0$. But $f(-c) \leq p(-c)$; hence $f(c) = -f(-c) \geq -p(-c) \geq 0$.

EXERCISE 4.5.2. Prove properties (2) and (3) above.

EXERCISE 4.5.3. Show that a generalized limit of a sequence $(c_n)_{n\in\mathbb{N}}$ satisfies

$$\liminf_{n\to\infty} c_n \leq \text{L. i. m.}_n\, c_n \leq \limsup_{n\to\infty} c_n.$$

Notice that the previous exercise implies that if $(c_n)_{n\in\mathbb{N}}$ is a convergent sequence, then $\text{L. i. m.}_n\, c_n = \lim_{n\to\infty} c_n$.

EXERCISE 4.5.4.

(i) Show that $p((c_n)_{n\in\mathbb{N}}) = \limsup_{n\to\infty} c_n$ defines a subadditive functional on $\ell^\infty(\mathbb{R})$.
(ii) Use the subadditive functional p defined above to construct a different notion of "generalized limit" and show that properties (1) and (3) hold.
(iii) Show that this notion of a "generalized limit" may be constructed in such a way that property (2) fails.

4.6. Adjoint Operators and Isometries of Normed Linear Spaces

Given a bounded linear operator T on a Banach space V, there is an associated bounded linear operator, T^*, on the dual space of V, called the adjoint of T. The adjoint operator plays an important role in our further studies of normed linear spaces and, in particular, in our study of their isometries.

DEFINITION 4.6.1. Suppose V is a Banach space over F ($= \mathbb{R}$ or $\mathbb{C}$ as usual) and suppose $T \in \mathcal{BL}(V)$. We define an operator T^* on V^* by $T^*(\alpha)(v) = \alpha(Tv)$, for $v \in V$ and $\alpha \in V^*$. The operator T^* is called the *adjoint* of T.

REMARK 4.6.2. It is obvious that $T^* \in \mathcal{BL}(V^*)$ since both α and T are linear and bounded.

EXERCISE 4.6.3. Here are some simple facts about adjoint operators that are easy to prove. Suppose $T_1, T_2 \in \mathcal{BL}(V)$. Show that:

(i) $(T_1 + T_2)^* = T_1^* + T_2^*$;
(ii) $(zT)^* = zT^*$ for all $z \in \mathbb{C}$;
(iii) if I is the identity on V and I^* is the identity on V^*, then $(I)^* = I^*$;
(iv) $(T_1T_2)^* = T_2^*T_1^*$.

Here is the basic theorem about adjoint operators.

THEOREM 4.6.4. *Let V be a Banach space and let $T \in \mathcal{BL}(V)$. If T^* is the adjoint of T, then $||T|| = ||T^*||$.*

Proof. For $\alpha \in V^*$, $||T^*(\alpha)|| = \sup_{||v||=1} |\alpha(T(v))| \leq ||T||||\alpha||$. So $||T^*|| \leq |T||$. Now, assume $||T|| = 1$. Then, for any ε with $0 < \varepsilon < 1$, there exists $v \in V$ with $||v|| = 1$ such that $||Tv|| > \varepsilon$. By Other Hahn-Banach (see Exercise 4.4.5), there exists $\alpha \in V^*$ with $||\alpha|| = 1$ such that $\alpha(Tv) = ||Tv||$. But then

$$\varepsilon < |\alpha(Tv)| = |T^*(\alpha)(v)| \leq ||T^*(\alpha)|| \leq ||T^*||,$$

so $||T^*|| \geq 1$.

Given a Banach space V, we have defined its dual space V^* as $\mathcal{BL}(V, F)$. The elements of V^* are often called *bounded linear functionals* on V. Of course, we can also consider the dual of V^*, denoted V^{**}, the space of bounded linear functionals on V^*. There is a natural injection $v \mapsto f_v$ from V to V^{**} defined by $f_v(\alpha) = \alpha(v)$ for $\alpha \in V^*$.

EXERCISE 4.6.5. Show that $\|f_v\| = \|v\|$.

DEFINITION 4.6.6. If the injection $v \mapsto f_v$ is a surjection, we say that the Banach space V is *reflexive*.

EXAMPLE 4.6.7. If $1 < p < \infty$, then $\ell^p(F)$ is reflexive.

EXERCISE 4.6.8. Show that $\ell^1(F)$ is not reflexive.

In Chapter 2, we defined isometries of metric spaces. These comprise a very important class of transformations, and we now consider isometries of normed linear spaces. For these spaces, the definition of isometry takes the following form.

DEFINITION 4.6.9. Let V, W be normed linear spaces over F. A surjection $T : V \longrightarrow W$ is an *isometry* if, for all $v, v' \in V$, $\|Tv - Tv'\| = \|v - v'\|$.

EXERCISE 4.6.10.

(i) If T is an isometry from V to W, show that T must be an injection.
(ii) If T is a linear isometry from V to W, show that $\|T\| = 1$.
(iii) If T is an isometry from V to W, show that T is a homeomorphism.

EXERCISE 4.6.11. Show that if V and W are finite-dimensional normed linear spaces over F of the same dimension, then V and W are homeomorphic. Thus, for example, $\ell_n^p(F)$ and $\ell_n^q(F)$ are homeomorphic for any $p, q \geq 1$.

The immediate question is, "Must isometries be linear maps?" Well, not quite, since translations in any normed linear space are isometries and translations do not take 0 to 0. That is, if V is a normed linear space and v_0 is a fixed element of V, the map $T : V \longrightarrow V$ defined by $T(v) = v + v_0$ is a *translation*. If $v_0 \neq 0$, T is not linear since $T(0) = v_0$, but it is clear that T is an isometry.

LEMMA 4.6.12. *Let V and W be normed linear spaces and let $T : V \longrightarrow W$ be an isometry. Then T can be modified by composition with a translation to obtain an isometry that sends 0 to 0.*

Proof. If $T(0) = w_0$, define $T'(v) = T(v) - w_0$. Then T' is an isometry, and $T'(0) = 0$.

If $F = \mathbb{C}$, then there are simple examples of isometries that take 0 to 0 but are not linear. For example, the map $\mathbb{C} \longrightarrow \mathbb{C}$ that sends z to $\bar{z}$ is norm-preserving but not $\mathbb{C}$-linear. More generally, in any of the spaces $\ell^p_n(\mathbb{C})$ or even $\ell^p(\mathbb{C})$, we can take a map that is the identity in some of the coordinates and complex conjugation in the others. This yields a map that sends 0 to 0 and is an isometry, but it is certainly not $\mathbb{C}$-linear.

However, for real normed linear spaces, we do have the following rather amazing theorem.

THEOREM 4.6.13 **(Mazur-Ulam)**. *Suppose that V and W are real normed linear spaces. Let $T : V \longrightarrow W$ be an isometry such that $T(0) = 0$. Then T is linear.*

Proof. Assume for the moment that V is just a metric space. Fix two points v_1, v_2 in V, and let

$$H_1 = \{v \in V |\, d(v, v_1) = d(v, v_2) = \frac{1}{2} d(v_1, v_2)\}.$$

Recall from Definition 2.3.38 that, for any nonempty subset A of a metric space X, the diameter is defined to be

$$\text{diam}(A) = \sup\{d(x, y) |\, x, y \in A\}.$$

Now, for each integer $n > 1$, define

$$H_n = \{v \in H_{n-1} |\, d(v, w) \leq \frac{1}{2}\text{diam}(H_{n-1}) \text{ for all } w \in H_{n-1}\}.$$

If the sets H_n are nonempty, then it is easy to see that they form a nested sequence of sets, $H_n \supseteq H_{n+1}$, and that $\text{diam}(H_n) \leq \frac{1}{2^{n-1}} d(v_1, v_2)$. Therefore, if $H_n \neq \varnothing$ for all n,

$$\lim_{n\to\infty} \text{diam}(H_n) = 0.$$

Thus, the intersection of all the H_n, if nonempty, is a single point that we call the *center* of v_1, v_2 and denote by c.

Now, consider the case when the metric space V is a normed linear space. Using the additive structure, we can define, for any $v \in V$, $\bar{v} = v_1 + v_2 - v$. Then, for each n, if $v \in H_n$, we have $\bar{v} \in H_n$ as well.

We claim that $c = \frac{1}{2}(v_1 + v_2)$ is the center of v_1, v_2. Clearly $c \in H_1$. Suppose $c \in H_{n-1}$. For any $v \in H_{n-1}$, we know that $\bar{v} \in H_{n-1}$, and since

$$2||c - v|| = ||v_1 + v_2 - 2v|| = ||\bar{v} - v|| \leq \text{diam}(H_{n-1}),$$

it follows that $c \in H_n$. By induction, c is the center of v_1, v_2.

Let T be an isometry from V to W that takes 0 to 0. Since the center depends only on the metric space structure, the center of any two points v_1, v_2 in V is mapped to the center of $T(v_1), T(v_2)$ in W. Thus,

$$T\left(\frac{1}{2}(v_1 + v_2)\right) = \frac{1}{2}(T(v_1) + T(v_2)).$$

When $v_2 = 0$, this implies that $T(\frac{1}{2}v) = \frac{1}{2}T(v)$ for all $v \in V$. This means that $T(\frac{1}{2}(v_1 + v_2)) = \frac{1}{2}T(v_1 + v_2) = \frac{1}{2}(T(v_1) + T(v_2))$. So T is additive. Using the standard techniques (see Section 1.7), we can now conclude that $T(rv) = rT(v)$ for any $r \in \mathbb{Q}$. Since isometries are continuous, this means that $T(\alpha v) = \alpha T(v)$ for all $\alpha \in \mathbb{R}$.

Remark 4.6.14. Observe that the proof of additivity above also works if the spaces V and W are complex normed linear spaces. Thus, for complex linear spaces, an isometry T is additive, but we just do not know whether T is complex linear in that case. We have seen above, from the simple example $T(z) = \bar{z}$, that when working with complex normed linear spaces, an isometry that sends 0 to 0 may not be linear.

The following example shows that the hypothesis of surjectivity is crucial in the Mazur-Ulam theorem.

Example 4.6.15. Let $V = (\mathbb{R}, |\cdot|)$ and let $W = \ell_2^2(\mathbb{R})$. Define $T : \mathbb{R} \longrightarrow \ell_2^2(\mathbb{R})$ by $T(x) = (x/\sqrt{2}, |x|/\sqrt{2})$. Then T is norm-preserving, and $T(0) = (0, 0)$. However, T is not linear.

Exercise 4.6.16. Find all isometries from $(\mathbb{R}, |\cdot|)$ to $\ell_2^2(\mathbb{R})$ that send 0 to $(0, 0)$.

It is important to have a name for the situation described in the Mazur-Ulam theorem.

Definition 4.6.17. Let V, W be normed linear spaces. A map $T : V \longrightarrow W$ is called an *isometric isomorphism* if T is both an isomorphism of vector spaces and an isometry. When such a T exists, the spaces are called *isometrically isomorphic*. This is one possible way to classify normed linear spaces, since isometric isomorphism is an equivalence relation on these spaces.

Exercise 4.6.18.

(i) Show that V is reflexive iff the map $v \mapsto f_v$ is an isometric isomorphism.
(ii) Find an example of a normed linear space V such that V and V^{**} are isometric but V is not reflexive.

4.7. Concrete Facts about Isometries of Normed Linear Spaces

We turn now to a more detailed study of isometries of particular normed linear spaces. For this section, all isometries of normed linear spaces will

be taken to be linear. In particular, they send 0 to 0 and preserve scalar multiplication. Suppose V and W are Banach spaces over F. The principal question is whether there are any isometries at all from V to W and, if there are, how to describe them. The first case to consider is the set

$$\mathcal{I}(V) = \{T \in \mathcal{BL}(V) | \, T \text{ is an isometry}\}.$$

It is clear that $\mathcal{I}(V)$ is a group.

DEFINITION 4.7.1. The group $\mathcal{I}(V)$ is called the *isometry group* of V.

EXERCISE 4.7.2. Suppose that V and W are finite-dimensional normed linear spaces over F that are isometric. Show that $\dim V = \dim W$.

EXAMPLES 4.7.3.

(i) Consider the spaces $\ell_n^p(F)$. There are some obvious isometries of these spaces. For example, any permutation of the coordinates is an isometry. Changing the sign of any number of coordinates is also an isometry. Since there are $n!$ permutations of the coordinates and 2^n sign changes in the coordinates, there are at least $n!2^n$ isometries of $\ell_n^p(F)$.

(ii) Consider the infinite-dimensional spaces $\ell^p(F)$. The isometries in (i) can be extended to these infinite-dimensional spaces. Needless to say, there are a very large number of isometries here.

(iii) Let X be a set, and let $V = \mathcal{B}(X, \mathbb{C})$. Let $f \in V$ be such that $|f(x)| = 1$ for all $x \in X$. Define $M_f : V \longrightarrow V$ by $(M_f g)(x) = f(x)g(x)$. Then M_f is an isometry of V.

In order to completely categorize the isometries of $\ell_n^p(F)$, we need the following theorem.

THEOREM 4.7.4. *Suppose that V is a Banach space and $T \in \mathcal{BL}(V)$. If T is an isometry, then the adjoint operator T^* is also an isometry.*

Proof. First observe that, in V, the set $\{v \mid \|v\| = 1\}$ is the same as the set $\{v \mid \|Tv\| = 1\}$ since T is an isometry. Now, for $\alpha \in V^*$, we have

$$\begin{aligned} \|T^*\alpha\| &= \sup_{\|v\|=1} |T^*\alpha(v)| \\ &= \sup_{\|v\|=1} |\alpha(Tv)| \\ &= \sup_{\|Tv\|=1} |\alpha(Tv)| \\ &= \sup_{\|w\|=1} |\alpha(w)| \\ &= \|\alpha\|. \end{aligned}$$

EXERCISE 4.7.5. (i) Finish the proof by showing that T^* is surjective.

(ii) Show that the converse of this theorem holds. That is, show that if T^* is an isometry, then T is also an isometry.

REMARK 4.7.6. Suppose that $1 < p < \infty$ and suppose that q is defined by the equation $\frac{1}{p} + \frac{1}{q} = 1$. Suppose that $V = \ell_n^p(F)$. If $T \in \mathcal{I}(V)$, we can represent T by an $n \times n$ matrix over $\mathbb{R}$ relative to the standard basis, $A = (a_{ij})_{i,j=1,\ldots,n}$.

EXERCISE 4.7.7. We observed above that permutations of coordinates and sign changes on coordinates give isometries of $\ell_n^p(F)$. Find the matrices that represent these isometries with respect to the standard basis.

DEFINITION 4.7.8. If $A = (a_{ij})_{i,j=1,\ldots,n}$ is an $n \times n$ matrix over any field F, the *transpose* of A is ${}^tA = (a_{ji})_{i,j=1,\ldots,n}$.

EXERCISE 4.7.9. Suppose that V is a finite-dimensional normed linear space over F. Take $T \in \mathcal{BL}(V)$ and represent it by a matrix A, relative to some basis. Show that T^* is represented by ${}^tA$ relative to the dual basis.

Now we can prove the following theorem.

THEOREM 4.7.10. *Suppose $1 < p < \infty$, $p \neq 2$, and $T \in \mathcal{I}(\ell_n^p(\mathbb{R}))$. Then if A is the matrix of T relative to the standard basis, then A is the product of a permutation matrix and a diagonal matrix with 1's or -1's on the diagonal. The group $\mathcal{I}(\ell_n^p(\mathbb{R}))$ is the group P_n discussed in the project in Section* B.6.2 *of Appendix* B.

Proof. Suppose $1 < p < 2$ and T is an isometry of $\ell_n^p(\mathbb{R})$ that is represented by a matrix $A = (a_{ij})_{i,j=1,\ldots,n}$ relative to the standard basis. Then T^* is an isometry of $\ell_n^q(\mathbb{R})$, where $\frac{1}{p} + \frac{1}{q} = 1$. The adjoint T^* is represented by the matrix ${}^tA = (a_{ji})_{i,j=1,\ldots,n}$ relative to the standard basis. Note that $2 < q < \infty$. Now, $Ae_j = (a_{1j}, a_{2j}, \ldots, a_{nj})$. Hence, for all j, $\sum_{i=1}^n |a_{ij}|^p = 1$. Similarly ${}^tAe_i = (a_{i1}, a_{i2}, \ldots, a_{in})$. Thus, for all i, we have $\sum_{j=1}^n |a_{ij}|^q = 1$. It follows that

$$\sum_{i,j=1}^n |a_{ij}|^p = n = \sum_{i,j=1}^n |a_{ij}|^q.$$

This implies that $|a_{ij}| \leq 1$ for all i, j and, since $p < q$, we have $|a_{ij}|^q \leq |a_{ij}|^p$ with equality iff $|a_{ij}| = 0$ or $|a_{ij}| = 1$. We conclude that $|a_{ij}| = 0$ or 1. This implies that each row and column of T contains exactly one nonzero element, and this element is ± 1.

EXERCISE 4.7.11.

(i) Determine the isometry group of $\ell_n^1(\mathbb{R})$.
(ii) Determine the isometry group of $\ell_n^\infty(\mathbb{R})$.

EXERCISE 4.7.12. Find $\mathcal{I}(\ell_n^p(\mathbb{C}))$ for $1 < p < \infty$, $p \neq 2$.

We now turn to the infinite-dimensional case.

LEMMA 4.7.13. *Let $p \in \mathbb{R}$, $1 \leq p < \infty$, and $p \neq 2$. Let $x = (x_j)_{j=1}^{\infty}, y = (y_j)_{j=1}^{\infty} \in \ell^p(\mathbb{C})$. Suppose that $\|x+y\|^p = \|x-y\|^p = \|x\|^p + \|y\|^p$, that is,*

$$\sum_{j=1}^{\infty} |x_j + y_j|^p = \sum_{j=1}^{\infty} |x_j - y_j|^p = \sum_{j=1}^{\infty} |x_j|^p + |y_j|^p.$$

Then, for all j, either $x_j = 0$ or $y_j = 0$.

Proof. We break the proof into two parts. First, let $p > 2$ and take $z, w \in \mathbb{C}$. Then

$$|z+w|^p = |(z+w)^2|^{p/2} = ((z+w)(\bar{z}+\bar{w}))^{p/2} = (|z|^2 + |w|^2 + 2\mathrm{Re} z\overline{w})^{p/2}.$$

Similarly, $|z-w|^p = (|z|^2 + |w|^2 - 2\mathrm{Re} z\overline{w})^{p/2}$. Now, since $p/2 > 1$, $(|z|^2 + |w|^2)^{p/2} \geq |z|^p + |w|^p$ with equality if and only if $z = 0$ or $w = 0$. This follows easily from the fact that if $a > 0$ and $t > 1$ are real numbers, then $(1+a)^t > 1 + a^t$. On the other hand, we have

$$(|z|^2 + |w|^2 + 2\mathrm{Re} z\overline{w})^{p/2} + (|z|^2 + |w|^2 - 2\mathrm{Re} z\overline{w})^{p/2} \geq 2(|z|^2 + |w|^2)^{p/2}.$$

This follows from the convexity of the function $f(a) = a^{p/2}$. Next, take the case $1 \leq p < 2$. Here, we have $|z|^p + |w|^p \geq \frac{1}{2}(|z+w|^p + |z-w|^p)$ with equality if and only if $z = 0$ or $w = 0$. This follows easily from the argument above since, if $a > 0$ and $0 < t < 1$, then $(1+a)^t < 1 + a^t$.

THEOREM 4.7.14. *Suppose $1 \leq p < \infty$, $p \neq 2$, and let T be an isometry of $\ell^p(\mathbb{R})$. Then T consists of a product of a permutation of the coordinates and sign changes in the coordinates.*

Proof. Since T is an isometry, for $k \neq j$, we have $\|Te_j + Te_k\|^p = \|Te_j - Te_k\|^p = \|Te_j\|^p + \|Te_k\|^p$. It follows from the lemma that Te_j and Te_k have disjoint support. Suppose that $Te_k = \sum_{j=1}^{\infty} \alpha_{kj} e_j$ and that $\alpha_{km} \neq 0$ and $\alpha_{kn} \neq 0$. Then, for any $l \neq k$ we have $\alpha_{lm} = \alpha_{ln} = 0$. Let v be any element of $\ell^p(\mathbb{R})$. We wish to show that it is not possible to have $Tv = e_n$. Write $v = v_k e_k + \sum_{l \neq k} v_l e_l$. Then $Tv = v_k Te_k + (\sum_{l \neq k} v_l Te_l)$. Consider the n-th and m-th coordinates of each term. For the first term, we know they are not zero and, from the lemma, we know that in the second term, they are both zero. Hence, regardless of the value of v_k, we cannot get e_n.

At this point the natural question to ask is whether there are isometries from $\ell_n^p(F)$ to $\ell_n^q(F)$ or $\ell^p(F)$ to $\ell^q(F)$ where $p \neq q$. The fact is that this practically never happens, as indicated by the following theorem and two challenge problems at the beginning of Chapter 5.

THEOREM 4.7.15. *There is an isometry between $\ell_2^1(\mathbb{R})$ and $\ell_2^\infty(\mathbb{R})$.*

Proof. To illustrate this, simply rotate the unit ball in $\ell_2^1(\mathbb{R})$ by 45 degrees and expand by a factor of $\sqrt{2}$. The corresponding matrix is

$$T = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

As stated in Challenge Problem 5.1, the only isometries between $\ell_n^p(\mathbb{R})$ and $\ell_n^q(\mathbb{R})$ happen in the case above when $n = 2$, $p = 1$, $q = \infty$.

Now consider the case $p = 2$. If $T \in \mathcal{I}(\ell_2^2(\mathbb{R}))$, then T can be represented by a 2×2 matrix

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

From the discussion in the project in Section B.6.3 of Appendix B, T can be written in the form

$$T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

or

$$T = \begin{pmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Thus, it follows that $\mathcal{I}(\ell_2^2(\mathbb{R}))$ is the orthogonal group $O_2(\mathbb{R})$.

THEOREM 4.7.16. *The isometry group of $\ell_2^2(\mathbb{R})$ is $O_2(\mathbb{R})$. The rotations of $\ell_2^2(\mathbb{R})$ form the subgroup $SO_2(\mathbb{R})$ of $O_2(\mathbb{R})$ made up of matrices of determinant* 1.

4.8. Locally Compact Groups

The group $GL_n(F)$ (where $F = \mathbb{R}$, $\mathbb{C}$, or $\mathbb{Q}_p$) is an example of a very important class of topological groups, that is, the locally compact groups.

DEFINITION 4.8.1. A topological group G is a *locally compact group* if every point in G has a neighborhood whose closure is compact.

EXERCISE 4.8.2.

(i) Show that $(\mathbb{R}^n, +)$, $(\mathbb{C}^n, +)$, $(\mathbb{Q}_p^n, +)$, $(\mathbb{R}^\times, \cdot)$, $(\mathbb{C}^\times, \cdot)$, $(\mathbb{Q}_p^\times, \cdot)$, and $(\mathbb{T}^n, \cdot)$, are locally compact groups with the usual topologies. The remaining examples require a little further discussion. Note that $\mathbb{T}^n$ is compact.

(ii) Show that $GL_n(F)$ is open and dense in $M_n(F)$.

(iii) Show that $GL_n(F)$ is a locally compact group.

(iv) Show that a closed subgroup of a locally compact group is a locally compact group.

Many important examples of topological groups are subgroups of $GL_n(F)$, where $F = \mathbb{R}, \mathbb{C}$, or $\mathbb{Q}_p$.

DEFINITION 4.8.3. Let $F = \mathbb{R}, \mathbb{C}$, or $\mathbb{Q}_p$. The following examples are closed subgroups of $GL_n(F)$ and hence are locally compact groups.

(a) The *special linear group* is defined by

$$SL_n(F) = \{g \in GL_n(F) \mid \det(g) = 1\}.$$

This is a closed subgroup since the determinant is a continuous function from $GL_n(F)$ to $F^\times$.

(b) The *orthogonal group* is defined by

$$O_n(F) = \{g \in GL_n(F) \mid g^t g = I\}.$$

Since $\det({}^t g) = \det(g)$, it follows that $\det(g^2) = 1$ and $\det(g) = \pm 1$.

(c) The *special orthogonal group* is defined by

$$SO_n(F) = \{g \in O_n(F) \mid \det(g) = 1\}.$$

REMARK 4.8.4. The groups $SL_n(F)$ and $SO_n(F)$ are called *classical groups* over F. In addition, there is one more collection of classical groups. These are called symplectic groups.

Let J be the $2n \times 2n$ matrix

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

where I_n is the $n \times n$ identity matrix.

DEFINITION 4.8.5. Let $F = \mathbb{R}, \mathbb{C}$, or $\mathbb{Q}_p$. The *symplectic group* $Sp_{2n}(F)$ is defined by

$$Sp_{2n}(F) = \{g \in GL_{2n}(F) \mid gJ^t g = J\}.$$

EXERCISE 4.8.6. Show that if $g \in Sp_{2n}(F)$, then $\det g = 1$.

We will not spend a great deal of time analyzing the structure of $Sp_{2n}(F)$. It is a locally compact, noncompact group.

Let us look at $O_n(\mathbb{R})$ and $SO_n(\mathbb{R})$. The condition $g^t g = I$ implies that ${}^t g = g^{-1}$. So we also have the condition ${}^t g g = I$. Putting these two together, we get the fact that, for any fixed i, $\sum_{j=1}^n a_{ij}^2 = 1$, and for any fixed j, $\sum_{i=1}^n a_{ij}^2 = 1$. It follows immediately that $O_n(\mathbb{R})$ is a closed bounded subset of $\mathbb{R}^{n^2}$ so we have proved the following:

THEOREM 4.8.7. *The group $O_n(\mathbb{R})$ (and consequently $SO_n(\mathbb{R})$) is compact.*

EXERCISE 4.8.8. Show that the groups $GL_n(F)$, $SL_n(F)$, $O_n(\mathbb{C})$, $SO_n(\mathbb{C})$, and $Sp_{2n}(\mathbb{C})$ are locally compact but not compact.

In the exercise above, we observed that $O_n(\mathbb{C})$ and $SO_n(\mathbb{C})$ are not compact. Is there an analog of the compact groups $O_n(\mathbb{R})$ and $SO_n(\mathbb{R})$ over the complex numbers? The answer is yes.

DEFINITION 4.8.9.

(a) The *unitary group* is defined by

$$U_n = \{g \in GL_n(\mathbb{C}) \mid g^t \bar{g} = I\}.$$

Here, $\bar{g}$ is the complex conjugate of g; that is, if $g = (a_{ij})_{i,j=1,\ldots,n}$, then $\bar{g} = (\overline{a_{ij}})_{i,j=1,\ldots,n}$.

(b) The *special unitary group* is defined by

$$SU_n = \{g \in U_n \mid \det(g) = 1\}.$$

If $g \in U_n$, $g = (a_{ij})_{i,j=1,\ldots,n}$, then for any fixed i, we have

$$\sum_{j=1}^{n} |a_{ij}|^2 = 1,$$

and, for any fixed j, we have

$$\sum_{i=1}^{n} |a_{ij}|^2 = 1.$$

Consequently, U_n is a closed bounded subset of $\mathbb{C}^{n^2}$.

THEOREM 4.8.10. *The group U_n (and hence SU_n) is compact.*

The groups $O_n(\mathbb{R})$, $SO_n(\mathbb{R})$, U_n, and SU_n are called *classical compact groups*. Along with these, we also have a compact symplectic group, which will not be part of our discussion here. The following facts about their structure are easy to prove with some elementary linear algebra.

THEOREM 4.8.11.

(i) *The center of U_n is the set of matrices*

$$Z = \left\{ \begin{pmatrix} e^{i\theta} & 0 & \ldots & 0 \\ 0 & e^{i\theta} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & e^{i\theta} \end{pmatrix} \right\}$$

where $0 \leq \theta < 2\pi$.

(ii) *Let T be the set of diagonal matrices in U_n, that is,*

$$T = \left\{ \left. \begin{pmatrix} e^{i\theta_1} & 0 & \ldots & 0 \\ 0 & e^{i\theta_2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & e^{i\theta_n} \end{pmatrix} \right| 0 \leq \theta_j < 2\pi \right\}.$$

Then

$$U_n = \{gtg^{-1} \mid g \in U_n \text{ and } t \in T\}.$$

(iii) *Every element of U_n can be written as the product of an element of SU_n and an element in the center of U_n.*

Proof. Statements (i) and (ii) follow easily from elementary properties of diagonalization of matrices over $\mathbb{C}$.

For (iii), suppose $g \in U_n$. Then $\det(g^t\overline{g}) = |\det(g)^2| = 1$. Thus $|\det(g)| = 1$. If $\det(g) = e^{i\theta}$, then the matrix

$$d = \begin{pmatrix} e^{i\theta/n} & 0 & \dots & 0 \\ 0 & e^{i\theta/n} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & e^{i\theta/n} \end{pmatrix}$$

is in Z, and the matrix $gd^{-1} = g' \in SU_n$. Consequently, we can write $g = dg'$.

EXERCISE 4.8.12. Find the intersection of SU_n and the center of U_n.

EXERCISE 4.8.13.

(i) Show that $SL_n(F)$ is a normal subgroup of $GL_n(F)$.
(ii) Show that $SO_n(F)$ is a normal subgroup of $O_n(F)$.
(iii) Show that SU_n is a normal subgroup of U_n.

The classical compact groups have an additional important property.

DEFINITION 4.8.14. A subgroup K of a locally compact group G is called a *maximal compact subgroup* if the following hold:

(a) K is a compact subgroup of G;
(b) if H is a compact subgroup of G containing K, then $H = K$.

EXERCISE 4.8.15. Show that the following compact groups are maximal compact subgroups in the indicated noncompact group:

(i) U_n in $GL_n(\mathbb{C})$;
(ii) SU_n in $SL_n(\mathbb{C})$;
(iii) $O_n(\mathbb{R})$ in $GL_n(\mathbb{R})$;
(iv) $SO_n(\mathbb{R})$ in $SL_n(\mathbb{R})$.

4.9. Hilbert Spaces

The geometry of Banach spaces becomes much more rigid when additional properties are added. When we discuss Hilbert spaces, we can think of a Banach space V in which the norm satisfies the *parallelogram law*, that is,

$$||v + w||^2 + ||v - w||^2 = 2(||v||^2 + ||w||^2)$$

for all $v, w \in V$. Since we have been studying Banach spaces, this might appear to be a good way to introduce Hilbert spaces. In fact, the norm in a Hilbert space arises from an inner product, and this is the most direct way to approach it.

DEFINITION 4.9.1. Suppose that V is a vector space over $\mathbb{C}$. A *positive definite Hermitian form* on V is a map

$$(\cdot|\cdot) : V \times V \longrightarrow \mathbb{C}$$

satisfying the following properties:

(a) for all $v \in V$, $(v|v) \geq 0$, and $(v|v) = 0$ if and only if $v = 0$,
(b) $(\alpha v_1 + v_2|w) = \alpha(v_1|w) + (v_2|w)$ for all $\alpha \in \mathbb{C}$, $v_1, v_2, w \in V$,
(c) $(v|w) = \overline{(w|v)}$ for all $v, w \in V$.

There are some elementary properties of Hermitian forms that follow immediately.

EXERCISE 4.9.2. Prove that, for all $\alpha \in \mathbb{C}$ and $v, w_1, w_2 \in V$, we have

$$(v|\alpha w_1 + w_2) = \bar{\alpha}(v|w_1) + (v|w_2).$$

Thus, we may say that a Hermitian form is linear in the first variable and *conjugate linear* in the second variable.

We now define a norm on V as follows.

DEFINITION 4.9.3. For all $v \in V$, set

$$||v|| = (v|v)^{\frac{1}{2}}.$$

Of course we have to prove that this defines a norm on V. To begin, it is certainly true from the definition that $\|v\| \geq 0$ and $\|v\| = 0$ iff $v = 0$. Moreover, for any $\alpha \in \mathbb{C}$,

$$\|\alpha v\| = (\alpha v|\alpha v)^{\frac{1}{2}} = (\alpha\bar{\alpha})^{\frac{1}{2}}(v|v)^{\frac{1}{2}} = |\alpha|\,||v||.$$

To prove the triangle inequality, we first prove the Cauchy-Schwarz inequality.

LEMMA 4.9.4 **(Cauchy-Schwarz inequality)**. *Suppose V is a vector space over $\mathbb{C}$ and let $(\cdot|\cdot)$ is a positive definite Hermitian form on V. Then, for $v, w \in V$, $|(v|w)| \leq \|v\|\,\|w\|$.*

Proof. If $(v|w) = 0$, the inequality is obvious. Otherwise, for any $\alpha \in \mathbb{C}$,

$$0 \leq (v - \alpha w|v - \alpha w) = \|v\|^2 - \ \alpha(w|v) - \bar{\alpha}(v|w) + |\alpha|^2\|w\|^2.$$

Setting $\alpha = \frac{\|v\|^2}{(w|v)}$, we get the Cauchy-Schwarz inequality.

COROLLARY 4.9.5. *Let V be a vector space over $\mathbb{C}$ and let $(\cdot|\cdot)$ be a positive definite Hermitian form on V. For any $v, w \in V$, we have $\|v+w\| \leq \|v\| + \|w\|$.*

Proof. This follows from the identity

$$\begin{aligned} ||v + w||^2 = (v + w|v + w) &= \|v\|^2 + \|w\|^2 + (v|w) + (w|v) \\ &= \|v\|^2 + \|w\|^2 + 2\mathrm{Re}(v|w). \end{aligned}$$

But $\mathrm{Re}(v|w) \leq \|v\|\|w\|$.

EXERCISE 4.9.6.

(i) Determine when equality holds in the Cauchy-Schwarz inequality.
(ii) Determine when equality holds in the triangle inequality.
(iii) Show that the inner product $(\cdot|\cdot) : V \times V \longrightarrow \mathbb{C}$ is continuous.

So that's it. If we have a vector space V over $\mathbb{C}$ with a positive definite Hermitian form, we can define a norm on V as above.

DEFINITION 4.9.7. If V is complete with respect to the norm defined above, we say that V is a complex *Hilbert space.*

EXAMPLE 4.9.8. Let $V = \ell^2(\mathbb{C})$. If $z = (z_n)_{n\in\mathbb{N}}$ and $w = (w_n)_{n\in\mathbb{N}}$, we set $(z|w) = \sum_{n=1}^{\infty} z_n\overline{w_n}$.

EXERCISE 4.9.9. Show that the sum in the previous definition converges.

EXERCISE 4.9.10.

(i) Suppose that V is a Hilbert space with inner product $(\cdot|\cdot)$. Prove the parallelogram law.
(ii) Suppose V is a complex Banach space with $\|\cdot\|$ satisfying the parallelogram law. Using the parallelogram law, construct a positive definite Hermitian form on V that yields this norm.

Recall that a convex set in a linear space over $\mathbb{R}$ or $\mathbb{C}$ is a subset S with the property that, for any two points $v_1, v_2 \in S$, the line segment

$$\{(1-t)v_1 + tv_2 \mid 0 \leq t \leq 1\}$$

is a subset of S. Convex sets play a very important role in the study of Hilbert spaces.

THEOREM 4.9.11. *Let V be a Hilbert space. A nonempty closed convex set $C \subseteq V$ contains a unique element of smallest norm.*

Proof. Let $d = \text{glb}\{\|v\| \,|\, v \in C\}$. Take a sequence $(v_n)_{n\in\mathbb{N}}$ in C such that $(\|v_n\|)_{n\in\mathbb{N}}$ converges to d. Then for any $n, m \in \mathbb{N}$, $\frac{v_n+v_m}{2}$ is in C, and $\|v_n + v_m\| \geq 2d$. But, by the parallelogram law, we know that $\|v_n - v_m\|^2 = 2(\|v_n\|^2 + \|v_m\|^2) - \|v_n + v_m\|^2$. Hence, $(v_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. If $v_0 = \lim_{n\to\infty} v_n$, then $\|v_0\| = d$.

EXERCISE 4.9.12. Prove that v_0 is unique.

EXAMPLE 4.9.13. The above theorem is decidedly false in Banach spaces that are not Hilbert spaces. For example, let $V = \ell_2^{\infty}(\mathbb{R})$. Take $C = \{(1, y)| \, -1 \leq y \leq 1\}$. If $(1, y) \in C$, then $||(1, y)||_{\infty} = \max(\{1, |y|\}) = 1$ for any $y \in [-1, 1]$. So, for every point $(1, y) \in C$, we have $||(1, y)||_{\infty} = 1$.

EXERCISE 4.9.14.

(i) Show that Theorem 4.9.11 is false for $V = \ell_n^p(\mathbb{R})$ where $p \neq 2$.
(ii) Show that Theorem 4.9.11 is false for $V = \mathcal{BC}(X, F)$ with X a metric space.

DEFINITION 4.9.15. Suppose that V is a Hilbert space and that B is a nonempty subset of V. We define

$$B^{\perp} = \{v \in V | \, (v|b) = 0 \text{ for all } b \in B\}.$$

The space $B^{\perp}$ is called the *orthogonal complement* of B in V. The symbol $B^{\perp}$ is usually read B *perp.*

PROPOSITION 4.9.16. *If B is a nonempty subset of a Hilbert space V, then $B^{\perp}$ is a closed subspace of V.*

Proof. It is clear that $B^{\perp}$ is a subspace. Now suppose $(v_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $B^{\perp}$ and write $v = \lim_{n\to\infty} v_n$. Then for each $w \in B$, we have $\lim_{n\to\infty}(v_n|w) = (v|w) = 0$ by the continuity of the inner product.

EXERCISE 4.9.17. Show that $(B^{\perp})^{\perp} = B$ iff B is a closed subspace of V.

THEOREM 4.9.18. *Suppose V is a Hilbert space and V_0 is a closed subspace of V. Then any element $v \in V$ can be written uniquely in the form $v = v_0 + v_0'$, where $v_0 \in V_0$ and $v_0' \in V_0^{\perp}$.*

Proof. Take $v \in V$. If $v \in V_0$, we are done. If $v \notin V_0$, consider the coset

$$v + V_0 = \{v + w | \, w \in V_0\}.$$

This coset is obviously a closed convex set. Let v_0' be the element of smallest norm in $v + V_0$. Then $v_0' + V_0 = v + V_0$ and, by minimality, for all $w \in V_0$ and all $\lambda \in \mathbb{C}$, we have $||v_0' + \lambda w||^2 \geq ||v_0'||^2$. By setting $\lambda = -\frac{(v_0'|w)}{||w||^2}$, we find that $-\frac{|(v_0'|w)|^2}{||w||^2} \geq 0$. Hence, $(v_0'|w) = 0$, and $v_0' \in V_0^{\perp}$.

EXERCISE 4.9.19. Prove that the decomposition above is unique.

DEFINITION 4.9.20. Suppose that V is a Hilbert space and V_0 is a closed subspace. The element v_0 in V_0 described in Theorem 4.9.18 is called the *projection of v on V_0*. We write this projection as $P(v) = v_0$.

The following facts are immediate.

FACTS 4.9.21.

(1) The projection P is in $\mathcal{BL}(V)$.
(2) If $V_0 \neq 0$, then the projection P has norm 1.
(3) The projection P is an *idempotent operator*; that is, $P^2 = P$.
(4) If I is the identity operator on V, then $I - P$ is the projection of V on $V_0^{\perp}$.

DEFINITION 4.9.22. Let V and W be Hilbert spaces. Define an inner product on the vector space $V \oplus W$ by the formula

$$((v_1, w_1)|(v_2, w_2)) = (v_1|v_2)_V + (w_1|w_2)_W$$

with the obvious notation.

EXERCISE 4.9.23. Show that this defines an inner product on $V \oplus W$ with respect to which $V \oplus W$ is a Hilbert space.

EXERCISE 4.9.24. Let V be a Hilbert space and let $V_0 \subseteq V$ be a closed subspace. Show that the map

$$\begin{array}{rcl} V_0 \oplus V_0^\perp & \longrightarrow & V \\ (v, v') & \longmapsto & v + v' \end{array}$$

is an isometric isomorphism and describe its inverse.

The next thing to investigate is the dual space of a Hilbert space. So, if V is a Hilbert space, we want to describe V^*, the space of all bounded linear functionals on V. The first thing to note is that if w is a fixed element of V, the map $T_w(v) = (v|w)$ defines an element of V^*. We wish to show that every element of V^* is of this form.

THEOREM 4.9.25. *Let $f \in V^*$. Then there exists $w \in V$ such that $f(v) = (v|w)$ for all $v \in V$.*

Proof. If $f \equiv 0$, take $w = 0$. Otherwise, let $V_0 = \ker f \subseteq V$. Then V_0 is a closed subspace of V. Take $z \in V_0^\perp$ with $z \neq 0$, and set

$$w = \frac{\overline{f(z)}}{||z||^2} z.$$

Then $f(w) = (w|w)$. Now, V/V_0 is one-dimensional, and every $v \in V$ can be written uniquely in the form $v = v_0 + \lambda w$, with $v_0 \in V_0$ and $\lambda \in \mathbb{C}$. Then

$$f(v) = f(v_0 + \lambda w) = \lambda f(w) = \lambda(w|w) = (v_0 + \lambda w|w) = (v|w).$$

This result is sometimes called the *Riesz Representation Theorem*. Thus, in the case of a Hilbert space V, the dual space V^* can be identified with V by the above theorem.

EXERCISE 4.9.26. Show that $\|T_w\| = \|w\|$.

Because of this identification, in the context of Hilbert spaces, the adjoint of a bounded linear operator is slightly different from the Banach space definition.

DEFINITION 4.9.27. Let V be a Hilbert space. For $T \in \mathcal{BL}(V)$, we define $T^* \in \mathcal{BL}(V)$ by $(Tv|w) = (v|T^*w)$ for $v, w \in V$. The operator T^* is called the *Hilbert space adjoint* of T.

REMARK 4.9.28. It is clear that $T^* \in \mathcal{BL}(V)$. It is also clear that properties (i), (iii), and (iv) of Exercise 4.6.3 are satisfied by the Hilbert space adjoint. For property (ii), we have that $(zT)^* = \bar{z}T^*$ for $z \in \mathbb{C}$.

EXAMPLE 4.9.29. Let $V = \ell^2(\mathbb{C})$, and let $b = (b_n)_{n\in\mathbb{N}} \in \ell^\infty(\mathbb{C})$. Define the operator of multiplication by b as in Example 4.2.3(i). That is, if $a = (a_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{C})$, then $T_b(a) = (a_n b_n)_{n\in\mathbb{N}}$. It is clear that $T_b^* = T_{\bar{b}}$.

EXERCISE 4.9.30. Suppose that V is a Hilbert space and that $T \in \mathcal{BL}(V)$.

(i) Show that $\|T\| = \|T^*\|$ without using the Hahn-Banach theorem.
(ii) If T is invertible, show that $(T^*)^{-1} = (T^{-1})^*$.

DEFINITION 4.9.31. Let V be a Hilbert space and let $T \in \mathcal{BL}(V)$. We say that T is a *unitary operator on* V if T is a bijection and $(Tv_1|Tv_2) = (v_1|v_2)$ for all $v_1, v_2 \in V$.

EXERCISE 4.9.32. Show that a unitary map on a Hilbert space is an isometry.

EXERCISE 4.9.33.

(i) Let V be a Hilbert space. Show that the collection $\mathcal{U}(V)$ of unitary operators on V is a group under composition of operators.
(ii) Let $V = \ell^2_n(\mathbb{C})$. Show that the unitary group U_n is isomorphic to $\mathcal{U}(V)$.

EXAMPLE 4.9.34. Let $a = (a_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{C})$ and $b = (e^{i\theta_n})_{n\in\mathbb{N}}$. Then $b \in \ell^\infty(\mathbb{C})$. As in Example 4.9.29, we set $T_b(a) = (a_n e^{i\theta_n})_{n\in\mathbb{N}}$. Then T_b is a unitary operator on $\ell^2(\mathbb{C})$.

DEFINITION 4.9.35. Let V be a Hilbert space. An operator T in $\mathcal{BL}(V)$ is *selfadjoint* if $T^* = T$.

Selfadjoint operators are very important in the study of Hilbert spaces.

EXAMPLE 4.9.36. Let $(a_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{C})$, and let $(b_n)_{n\in\mathbb{N}} \in \ell^\infty(\mathbb{C})$. The operator T_b defined above is selfadjoint iff $b_n \in \mathbb{R}$ for all $n \in \mathbb{N}$.

The following two theorems relate projections to selfadjoint operators.

THEOREM 4.9.37. *Let V be a Hilbert space and $V_0 \neq 0$ a closed subspace of V. If P is the projection of V on V_0, then P is selfadjoint.*

Proof. Clear.

THEOREM 4.9.38. *Let V be a Hilbert space, and let $T \in \mathcal{BL}(V)$. If T is selfadjoint and $T^2 = T$, then T is a projection.*

Proof. Let $V_0 = T(V)$. Then $V_0 = \{v \in V \mid Tv = v\}$, so V_0 is closed. If $v_0' \in V_0^\perp$, then

$$(Tv_0'|Tv_0') = (v_0'|T^2v_0') = (v_0'|Tv_0') = 0,$$

since $Tv_0' \in V_0$. Hence, $Tv_0' = 0$, and, if $v = v_0 + v_0'$,

$$Tv = Tv_0 + Tv_0' = Tv_0 = v_0.$$

Thus, T is the projection on V_0.

EXERCISE 4.9.39. Let V be a Hilbert space and V_1, V_2 closed subspaces of V with corresponding projections P_1, P_2.

(i) Show that P_1P_2 is a projection iff P_1 and P_2 commute.

(ii) Show that $P_1 + P_2$ is a projection iff $P_1P_2 = 0$.
(iii) Show that $P_1 - P_2$ is a projection iff $P_1P_2 = P_2$.
(iv) Suppose each of P_1P_2, $P_1 + P_2$, and $P_1 - P_2$ is a projection. Find their images.

The most wonderful thing about Hilbert spaces is the fact that they have a topological basis consisting of vectors that have norm one and are pairwise orthogonal. This is what the inner product does for us.

DEFINITION 4.9.40. Let V be a Hilbert space, and let $S = \{v_i\}_{i\in I}$ be a set of nonzero vectors in V. The set S is an *orthogonal set* if, for any pair v_i, v_j, $i \neq j$, we have $(v_i|v_j) = 0$. The set S is an *orthonormal set* if S is an orthogonal set and $||v_i|| = 1$ for all $i \in I$.

EXERCISE 4.9.41. Suppose that S is an orthonormal set in a Hilbert space V. Show that S is linearly independent.

EXAMPLE 4.9.42. Consider the Hilbert space $V = \ell^2(\mathbb{C})$. If e_j is the standard basis vector with 1 in the j-th spot and 0 elsewhere, then $S = \{e_j\}_{j\in\mathbb{N}}$ is an orthonormal set in $\ell^2(\mathbb{C})$. As we have seen before, S is not an algebraic basis for V. However, it is a topological basis.

Now, the idea is, given a Hilbert space V, to find a topological basis that is orthonormal. We begin with some easy lemmas.

LEMMA 4.9.43 **(Bessel's inequality)**. *Let V be a Hilbert space, and let $S = \{v_i\}_{i\in I}$ be an orthonormal set in V. If $v \in V$ and $v_1, \ldots, v_n$ are elements of S, then*

$$\sum_{j=1}^{n} |(v|v_j)|^2 \leq ||v||^2.$$

Proof. Let $\alpha_j = (v|v_j)$. Then $0 \leq ||v - \sum_{j=1}^n \alpha_j v_j||^2$. Since $(v_i|v_j) = 0$ for $i \neq j$ and since $||v_j|| = 1$, by expanding the inner product, we get $||v||^2 - \sum_{j=1}^n |\alpha_j|^2 \geq 0$, which is the desired result.

COROLLARY 4.9.44. *The number of indices $i \in I$ for which $(v|v_i) \neq 0$ is finite or countable.*

Proof. For $n \in \mathbb{N}$, consider the number of elements $v_j \in S$ such that $|(v|v_j)| \geq \frac{1}{n}$. This is a finite set.

REMARK 4.9.45. It should be noted generally that there is no such thing as convergence for sums of uncountable sets of nonzero elements in $\mathbb{R}$.

COROLLARY 4.9.46. *If $v, w \in V$, then $\sum_{i\in I} |(v|v_i)\overline{(w|v_i)}| \leq ||v||\,||w||$, where $S = \{v_i\}_{i\in I}$ is an orthonormal set as above.*

Proof. Cauchy-Schwarz.

REMARK 4.9.47. There are many examples of orthonormal sets in analysis. We will encounter some of them in Chapter 7.

EXERCISE 4.9.48. Suppose $\{v_i\}_{i\in I}$ is an orthonormal set in a Hilbert space V. Show that if $i \neq j$, then $||v_i - v_j|| = \sqrt{2}$.

THEOREM 4.9.49. *Suppose V is a separable Hilbert space and S is an orthonormal set in V. Then S is either finite or countably infinite.*

Proof. Let $S = \{v_i\}_{i\in I}$. Now take $W = \{w_n\}_{n\in\mathbb{N}}$ to be a countable dense subset of V. If $v_j \in S$, there exists $w_m \in W$ such that $||v_j - w_m|| < \frac{\sqrt{2}}{3}$. For $v_k \in S$ with $k \neq j$, choose $w_h \in W$ such that $||v_k - w_h|| < \frac{\sqrt{2}}{3}$. Then from Exercise 4.9.48,

$$\sqrt{2} = ||v_k - v_j|| \leq ||v_k - w_h|| + ||v_j - w_m|| + ||w_h - w_m|| < 2\frac{\sqrt{2}}{3} + ||w_h - w_m||.$$

Hence, $||w_h - w_m|| \geq \frac{\sqrt{2}}{3}$, and so $h \neq m$. Consequently, S is in one-to-one correspondence with a subset of W.

THEOREM 4.9.50. *Let V be a Hilbert space, and let $S = \{v_i\}_{i\in I}$ be a countably infinite orthonormal set in V. Then the series $\sum_{j=1}^{\infty} \alpha_j v_j$ converges iff $\sum_{j=1}^{\infty} |\alpha_j|^2 < \infty$.*

Proof. Let $s_n = \sum_{j=1}^{n} \alpha_j v_j$. Then, for $m < n$, $||s_n - s_m||^2 = \sum_{j=m+1}^{n} |\alpha_j|^2$. Since V is complete, $(s_n)_{n\in\mathbb{N}}$ is Cauchy and hence convergent iff $\sum_{j=1}^{\infty} |\alpha_j|^2 < \infty$.

COROLLARY 4.9.51. *Let V be a Hilbert space, and let $\{v_i\}_{i\in I}$ be a countably infinite orthonormal set in V. If $v = \sum_{j=1}^{\infty} \alpha_j v_j$, then $\alpha_j = (v|v_j)$.*

Proof. Exercise.

THEOREM 4.9.52. *Suppose that V is a separable Hilbert space. Then V has a countable orthonormal basis; that is, there exists a countable orthonormal set in V whose closed linear span is V.*

Proof. There certainly exist orthonormal sets in V. Let S be a maximal orthonormal set of elements of V (Hausdorff maximality). Let V_0 be the closed linear span of S. If $V_0 \neq V$, then $V_0^{\perp}$ contains a unit vector orthogonal to S. This is a contradiction.

EXERCISE 4.9.53. Suppose that V is not separable. Is the theorem still true? Can you use the same proof?

If you like a hands-on approach and do not like Hausdorff maximality, try this.

THEOREM 4.9.54 **(Gram-Schmidt orthogonalization)**. *Let $S_0 = \{w_n\}$ be a finite or countable linearly independent set in a Hilbert space V. Then there exists an orthonormal set S having the same cardinal number and generating the same subspace as S_0.*

Proof. Let $v_1 = w_1$ and $u_1 = \frac{v_1}{||v_1||}$. Let $v_2 = w_2 - (w_2|u_1)u_1$ and $u_2 = \frac{v_2}{||v_2||}$. Continuing inductively, we set

$$v_{n+1} = w_{n+1} - \sum_{k=1}^{n} (w_{n+1}|u_k)u_k$$

and $u_{n+1} = \frac{v_{n+1}}{||v_{n+1}||}$. It is easy to check that the set $S = \{u_n\}$ has the required properties.

In this book, separable Hilbert spaces are the ones of most interest. The simplest example of an infinite-dimensional separable Hilbert space is $\ell^2(\mathbb{C})$. We could also use $\ell^2(\mathbb{R})$, but we are concentrating mainly on complex Hilbert spaces. You might think that we are really limited in our number of examples of separable Hilbert spaces. However, the following theorem tells us that life is not so bad after all.

THEOREM 4.9.55. *Any infinite-dimensional separable Hilbert space V is isometrically isomorphic to $\ell^2(\mathbb{C})$.*

Proof. Let $\{v_n\}_{n\in\mathbb{N}}$ be an orthonormal basis for V, and let $\{e_n\}_{n\in\mathbb{N}}$ be the usual basis for $\ell^2(\mathbb{C})$. Define an isometric isomorphism from V to $\ell^2(\mathbb{C})$ by sending v_n to e_n, extending by linearity and by continuity.

We shall see more examples of separable Hilbert spaces in Chapter 7.

4.10. Convergence and Selfadjoint Operators

There are three types of convergence of bounded linear operators on Hilbert spaces.

DEFINITION 4.10.1. Let $\{T_n\}$ be a sequence of bounded linear operators on a Hilbert space V.

(a) We say that $\{T_n\}$ *converges uniformly* to a bounded linear operator $T \in \mathcal{BL}(V)$ if $\|T_n - T\| \to 0$ as $n \to \infty$.
(b) We say that $\{T_n\}$ *converges strongly* to $T \in \mathcal{BL}(V)$ if, for each $v \in V$, $T_n v \to Tv$ as $n \to \infty$.
(c) We say that $\{T_n\}$ *converges weakly* to $T \in \mathcal{BL}(V)$ if, for any pair $v, w \in V$, $(T_n v|w) \to (Tv|w)$ as $n \to \infty$.

EXERCISE 4.10.2.

(i) Prove that uniform convergence implies strong convergence implies weak convergence.
(ii) Find examples to show that the reverse implications are false.

EXERCISE 4.10.3 **(Elementary properties of selfadjoint operators)**.

(i) Show that weak, strong, and uniform limits of selfadjoint operators are selfadjoint.

(ii) Show that the class of selfadjoint operators is closed under addition and multiplication by real scalars.
(iii) Show that if T_1 and T_2 are selfadjoint, then T_1T_2 is selfadjoint iff $T_1T_2 = T_2T_1$.
(iv) Show that T is selfadjoint iff $(Tv|v) \in \mathbb{R}$ for all $v \in V$.
(v) Show that if T is selfadjoint, then T^n is selfadjoint for all integers $n \geq 1$ and $\|T^n\| = \|T\|^n$.

For any selfadjoint operator T on a Hilbert space V, we define

$$m = \inf_{\|v\|=1} (Tv|v) \quad \text{and} \quad M = \sup_{\|v\|=1} (Tv|v).$$

EXERCISE 4.10.4. Show that if T is selfadjoint, then $\|T\| = \max(|m|, |M|)$.

DEFINITION 4.10.5. Let T be selfadjoint. We say that T is *positive* if $(Tv|v) \geq 0$ for all $v \in V$. We write $T \geq 0$ if T is positive and $T_1 \geq T_2$ if $T_1 - T_2$ is positive, for T_1 and T_2 selfadjoint.

Note that if T is positive, then $\|T\| = M$ in the notation above.

EXAMPLE 4.10.6. The selfadjoint operator T_b of Example 4.9.36 is positive iff $b_n \geq 0$ for all $n \in \mathbb{N}$.

THEOREM 4.10.7. *If $T \geq 0$, then there exists $S \geq 0$ such that $S^2 = T$.*

Proof. If $T = 0$, the result is obvious, so we may assume that $\|T\| = 1$. For any real number x with $|x| < 1$, we have

$$(1 - x)^{1/2} = 1 - c_1x - c_2x^2 - \cdots$$

where each $c_j > 0$ and $\sum_{j=1}^{\infty} c_j = 1$. In particular, the above expansion holds for $x = 1$ as well. Since $\|T\| = 1$, we have $0 \leq I - T \leq I$. Define T_n by

$$T_n = I - \sum_{j=1}^{n} c_j(I - T)^j.$$

Note that T_n is selfadjoint and, for $n \geq m$, we have $\|T_n - T_m\| \leq \sum_{j=m+1}^{\infty} c_j$. Thus, by Theorem 4.2.6, the sequence $(T_n)_{n\in\mathbb{N}}$ converges uniformly to a selfadjoint operator S. We have

$$S^2 = I - (I - T) = T$$

and, since $I - T \leq I$, the operator S is positive.

EXERCISE 4.10.8. Verify the claim that $S^2 = T$ in the above proof.

EXERCISE 4.10.9. Show that there is a *unique* S satisfying the conclusion of the above theorem.

REMARK 4.10.10. Observe that if T is selfadjoint, then $T^2 \geq 0$.

COROLLARY 4.10.11. *If T_1 and T_2 are commuting positive operators, then T_1T_2 is positive.*

Proof. We know that T_1T_2 is selfadjoint by Exercise 4.10.3. Let S_1 and S_2 be the unique positive square roots of T_1 and T_2, respectively. By the direct construction in the proof of Theorem 4.10.7, it is clear that S_1 and S_2 commute, since they are uniform limits of commuting operators. Thus by Exercise 4.10.3, S_1S_2 is selfadjoint and $(S_1S_2)^2 = T_1T_2$, and the result follows from Remark 4.10.10.

For a positive selfadjoint operator T, we denote by $T^{1/2}$ the unique positive square root of T, and we set $|T| = (T^2)^{1/2}$. Further, for T selfadjoint we define operators T^+ and T^- by

$$T^+ = \frac{1}{2}(|T| + T) \quad \text{and} \quad T^- = \frac{1}{2}(|T| - T).$$

EXERCISE 4.10.12. Let T be a selfadjoint operator.

(i) Show that T^+ and T^- are positive selfadjoint operators with $T = T^+ - T^-$ and $T^+T^- = 0$.
(ii) Show that this is the unique decomposition of T into the difference of positive commuting operators whose product is zero.

4.11. Independent Projects

The goal of the independent projects for this chapter is to prove the spectral theorem for compact selfadjoint operators on a Hilbert space. This is broken into three parts. First we introduce the spectrum of a selfadjoint operator. Second, we discuss compact operators on a Banach space. Finally, we present the spectral theorem.

4.11.1. The Resolvent Set and the Spectrum.

We are now going to consider the *spectrum* of a bounded linear operator on a Banach space V. First we present an exercise and then a lemma about operators and their inverses.

EXERCISE 4.11.1. Suppose that T_1 and T_2 are commuting bounded linear operators on V such that T_1T_2 is invertible. Show that each of T_1 and T_2 is invertible.

LEMMA 4.11.2. *If $\|I - T\| < 1$, then T is invertible.*

Proof. The series

$$\sum_{n=0}^{\infty} (I - T)^n$$

converges uniformly to T^{-1}.

COROLLARY 4.11.3. *If T is invertible, commutes with S, and $\|S - T\| < \|T^{-1}\|^{-1}$, then S is invertible as well.*

REMARK 4.11.4. In particular, this lemma implies that if $\|T\| < 1$, then $I - T$ is invertible.

DEFINITION 4.11.5. Let T be a bounded linear operator on the Banach space V. The *resolvent set* of T is

$$\rho(T) = \{\lambda \in \mathbb{C} \mid \lambda I - T \text{ is invertible}\}.$$

The *spectrum* of T is

$$\sigma(T) = \mathbb{C} \setminus \rho(T).$$

LEMMA 4.11.6. *We have* $\sigma(T) \subseteq \{\lambda \mid |\lambda| \leq \|T\|\}$.

Proof. If $|\lambda| > \|T\|$, then $\|\lambda^{-1}T\| < 1$, so by the above remark, $I - \lambda^{-1}T$ is invertible. But then $\lambda I - T$ is invertible as well, and $\lambda \notin \sigma(T)$.

THEOREM 4.11.7. *The spectrum* $\sigma(T)$ *is compact.*

Proof. By Corollary 4.11.3, the resolvent $\rho(T)$ is open, and hence $\sigma(T)$ is closed. By the previous lemma, $\sigma(T)$ is bounded, and the result follows.

DEFINITION 4.11.8. Let T be a bounded linear operator on the Banach space V. The complex number λ is an *eigenvalue* of T if there exists $v \in V$ with $v \neq 0$ and $Tv = \lambda v$. The *eigenspace* associated to the eigenvalue λ is the set

$$\{v \in V \mid Tv = \lambda v\}.$$

REMARK 4.11.9.

(1) The eigenspace associated to any eigenvalue is a closed subspace of V.
(2) The set of eigenvalues of T is contained in $\sigma(T)$.

EXERCISE 4.11.10. Let T be a selfadjoint operator on a Hilbert space V.

(i) Show that all eigenvalues of T are real.
(ii) Show that eigenvectors corresponding to distinct eigenvalues are orthogonal.

THEOREM 4.11.11. *Let* T *be a selfadjoint operator on a Hilbert space* V. *Then* $\sigma(T) \subseteq \mathbb{R}$.

Proof. We present this proof as an exercise in three parts.

EXERCISE 4.11.12.

(i) Show that if $\lambda \notin \mathbb{R}$, then $\lambda I - T$ is injective.
(ii) Show that if $\lambda \notin \mathbb{R}$, then $\lambda I - T$ has dense image in V.
(iii) Finally, show that if $\lambda \notin \mathbb{R}$, then $\lambda I - T$ is surjective, and complete the proof.

COROLLARY 4.11.13. *The spectrum of a selfadjoint operator* T *satisfies* $\sigma(T) \subseteq [-\|T\|, \|T\|]$.

THEOREM 4.11.14. *Let* T *be a selfadjoint operator on a Hilbert space* V. *Then* $\sigma(T) \subseteq [m, M]$.

Proof. We know that the spectrum lies in $\mathbb{R}$. Take $\lambda \in \mathbb{R} \setminus [m, M]$ and let d be the distance from λ to $[m, M]$.

EXERCISE 4.11.15. Show that $\|(T - \lambda I)v\| \geq d\|v\|$ for all $v \in V$, and conclude that T is one-to-one when restricted to its image.

EXERCISE 4.11.16. Show that $T - \lambda I$ is onto.

It follows that $\lambda \in \rho(T)$, which finishes the proof of the theorem.

THEOREM 4.11.17. *Both m and M lie in $\sigma(T)$.*

Proof. We break up the proof into exercises.

EXERCISE 4.11.18. Show that $((T - m)v|v) \geq 0$ for all $v \in V$.

EXERCISE 4.11.19. Show that $\|(T-m)v\|^4 \leq ((T-m)v|v)\|T-m\|^3\|v\|^2$ for all $v \in V$.

EXERCISE 4.11.20. Finish the argument by showing that $T-mI$ cannot be invertible and adapting these exercises to $T - MI$.

4.11.2. Compact Operators.

In this project, we treat a special class of bounded linear operators on Banach spaces. These are called *compact operators* and play a very important role in applications of operator theory, especially to partial differential equations.

DEFINITION 4.11.21. Let V and W be a Banach spaces and let $T : V \longrightarrow W$ be a bounded linear operator. We say that T is *compact* if, for every bounded sequence $(v_n)_{n\in\mathbb{N}}$ in V, the sequence $(Tv_n)_{n\in\mathbb{N}}$ contains a convergent subsequence in W.

EXERCISE 4.11.22. Show that the hypothesis of boundedness may be omitted in the above definition in that it follows from the convergence condition.

REMARK 4.11.23.

(1) If V is finite-dimensional, then any $T \in \mathcal{BL}(V)$ is compact.
(2) If V is a Hilbert space, V_0 is a closed subspace of V, and P is the projection of V onto V_0, then P is compact if and only if V_0 is finite-dimensional.
(3) More generally, any bounded linear operator T with finite-dimensional image (*finite rank*) is compact.
(4) The identity operator $V \longrightarrow V$ is compact if and only if V is finite-dimensional (see Exercise 4.11.29).
(5) Unitary operators on infinite-dimensional Hilbert spaces are not compact.

EXERCISE 4.11.24. Let V be a Banach space and $T \in \mathcal{BL}(V)$. Show that T is compact if and only if, for all bounded subsets $A \subseteq V$, the closure $\overline{T(A)}$ is compact.

EXERCISE 4.11.25. Let V be a Banach space and let $S, T \in \mathcal{BL}(V)$ with T compact. Show that ST and TS are compact.

THEOREM 4.11.26. *The subspace of compact operators is closed in $\mathcal{BL}(V)$ in the uniform topology.*

Proof. Let $(S_n)_{n\in\mathbb{N}}$ be a uniformly convergent sequence of compact operators converging to $S \in \mathcal{BL}(V)$, and let $(v_n)_{n\in\mathbb{N}}$ be a bounded sequence in V. The sequence $(S_1 v_n)_{n\in\mathbb{N}}$ contains a convergent subsequence $(v_{1n})_{n\in\mathbb{N}}$. The sequence $(v_{1n})_{n\in\mathbb{N}}$ is bounded, so $(S_2 v_{1n})_{n\in\mathbb{N}}$ contains a convergent subsequence $(v_{2n})_{n\in\mathbb{N}}$. Proceeding in this fashion, we obtain sequences $(v_{kn})_{n\in\mathbb{N}}$ for each positive integer k.

EXERCISE 4.11.27. Finish the proof by showing that $(Sv_{nn})_{n\in\mathbb{N}}$ converges.

LEMMA 4.11.28 **(Riesz's lemma)**. *Let V be a normed linear space and let W be a proper closed subspace of V. Then, for every $\varepsilon > 0$, there exists $u \in V$ such that $\|u\| = 1$ and the distance from u to W is at least $1 - \varepsilon$.*

Proof. Pick $v \in V \setminus W$. Since W is closed in V, the distance d from v to W is positive. Choose $w \in W$ such that

$$d \leq \|v - w\| \leq \frac{d}{1-\varepsilon}$$

and define

$$u = \frac{v - w}{\|v - w\|}.$$

Then, u is of norm 1, and, for all $w' \in W$, we have

$$\|u - w'\| = \left\| \frac{v-w}{\|v-w\|} - w' \right\| = \frac{\|v - (w + w'\|v - w\|)\|}{\|v - w\|} \geq \frac{d}{\|v-w\|} \geq 1 - \varepsilon,$$

which proves the claim.

EXERCISE 4.11.29. Show that the unit ball in a Banach space V is compact if and only if V is finite-dimensional.

THEOREM 4.11.30. *Let V be a Banach space and let $T \in \mathcal{BL}(V)$ be compact. If λ is a nonzero eigenvalue of T, then the eigenspace*

$$\{v \in V \mid Tv = \lambda v\}$$

is finite-dimensional.

Proof. Suppose that this eigenspace is infinite-dimensional. By Riesz's lemma, we can choose a sequence of eigenvectors $(v_n)_{n\in\mathbb{N}}$ such that $\|v_n\| = 1$ for all n and $\|v_n - v_m\| > 1/2$ for $n \neq m$. But then

$$\|Tv_n - Tv_m\| = |\lambda|\|v_n - v_m\| > \frac{1}{2}|\lambda|,$$

and $(Tv_n)_{n\in\mathbb{N}}$ cannot contain a convergent subsequence.

4.11.3. The Spectral Theorem for Compact Selfadjoint Operators.

THEOREM 4.11.31 **(Spectral Theorem for Compact Selfadjoint Operators)**. *Let V be a Hilbert space and let T be a compact selfadjoint operator on V. Then V has an orthonormal basis consisting of eigenvectors of T.*

The following theorem is the real workhorse used to prove the Spectral Theorem.

THEOREM 4.11.32. *Let V be a Hilbert space and let T be a compact selfadjoint operator on V. Then either $\|T\|$ or $-\|T\|$ is an eigenvalue of T, and there exists a corresponding eigenvector v such that $\|v\| = 1$ and $|(Tv|v)| = \|T\|$.*

Proof. Note that the existence of such a vector v follows easily from the statement that either $\|T\|$ or $-\|T\|$ is an eigenvalue.

By Exercise 4.10.4 and Theorem 4.11.11 there exists $\lambda = \pm\|T\|$ and a sequence of unit vectors $(v_n)_{n\in\mathbb{N}}$ in V such that $(Tv_n|v_n) \to \lambda$. But

$$\|Tv_n - \lambda v_n\|^2 \leq \|T\|^2 - 2\lambda(Tv_n|v_n) + \lambda^2$$

and so $\|Tv_n - \lambda v_n\| \to 0$. Since T is compact, $(Tv_n)_{n\in\mathbb{N}}$ contains a convergent subsequence, say $(w_k)_{k\in\mathbb{N}}$. But then $\|Tw_k - \lambda w_k\| \to 0$ implies that $(Tw_k)_{k\in\mathbb{N}}$ converges, since $\lambda \neq 0$. Suppose $w_k \to v$.

EXERCISE 4.11.33. Finish the proof by showing that $\|v\| = 1$ and that $Tv = \lambda v$.

The tools required to prove the the Spectral Theorem are now available. Here is a brief outline. Pick an eigenvector $v_1 \in V_1 = V$ as guaranteed by the previous theorem. Let V_2 be the orthogonal complement of $\mathbb{C}v_1$.

EXERCISE 4.11.34. Finish the proof of the Spectral Theorem by iterating the above construction.

Chapter 5

Differentiation

> The actual state of [infinitesimal calculus] at German universities still shows the same colorful diversity. On the one hand, there is the rigorous approach that begins with a six weeks long Dedekind treatment, and then deduces as natural and necessary consequences the concrete rules of differentiation and integration from the general concepts of number and function. On the other, there is the intuitive treatment that lets reign the magic of the differential, and even in the last hour of a two semester course does not dissipate by the sunshine of a clear conception of limit, the fog surrounding the indivisibles. And, in between, there are the shadings of hundreds of diagonals that one can construct between the two perpendicular directions of thought.
>
> – *Otto Toeplitz*
> *Jahresbericht der Deutschen Mathematiker (1927)*

Challenge Problems: V

PROBLEM 5.1. If $1 \leq p < q \leq \infty$, show that the only isometries between $\ell_n^p(\mathbb{R})$ and $\ell_n^q(\mathbb{R})$ for $n \geq 2$ occur when $n = 2$, $p = 1$, and $q = \infty$.

PROBLEM 5.2. If $1 \leq p < q \leq \infty$, determine all isometries between $\ell_n^p(\mathbb{C})$ and $\ell_n^q(\mathbb{C})$ for $n \geq 2$.

PROBLEM 5.3. If $1 \leq p < q \leq \infty$, find all isometries between $\ell^p(\mathbb{R})$ and $\ell^q(\mathbb{R})$. Do the same for $\ell^p(\mathbb{C})$ and $\ell^q(\mathbb{C})$.

PROBLEM 5.4. Let $A : \ell_n^p(\mathbb{R}) \longrightarrow \ell_n^q(\mathbb{R})$ be a bounded linear operator. If the matrix for A relative to the standard basis is (a_{ij}), find the norm of the operator in terms of the matrix entries a_{ij}.

PROBLEM 5.5. Show that every element g of $GL_n(\mathbb{C})$ can be written uniquely in the form $g = kau$ where $k \in U_n$, $a \in A^+$, the set of diagonal matrices with positive real entries, and $u \in N$ with complex entries. This is called the *Iwasawa decomposition* for $GL_n(\mathbb{C})$. Determine a similar decomposition for $SL_n(\mathbb{C})$, $GL_n(\mathbb{R})$, and $SL_n(\mathbb{R})$ (the reader should also consult the independent project in Section B.6.1 of Appendix B).

5.1. Review of Differentiation in One Variable

We assume that the reader is familiar with the standard properties of the derivative in one variable, and we will not review the computational aspects of elementary calculus. However, we shall establish rigorously those properties of the derivative in one variable that stem from the completeness of the real numbers. Many of the aspects of differentiation which occur in several variables are motivated by, and rely on, results in one variable.

We begin by defining the derivative of a real-valued function of one variable at a point.

DEFINITION 5.1.1. Let (a, b) be an interval in $\mathbb{R}$ and consider $f : (a, b) \longrightarrow \mathbb{R}$. We say that f is *differentiable* at a point $x \in (a, b)$ if there exists $L \in \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = L.$$

Observe that this definition can be rephrased in the following way. The function f is differentiable at $x \in (a, b)$ if there exists $L \in \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - Lh}{h} = 0.$$

The number L is called the *derivative* of f at x and is denoted by $f'(x)$ or $Df(x)$.

EXERCISE 5.1.2. If L exists, show that it is unique.

EXERCISE 5.1.3. Show that f is differentiable at $x \in (a, b)$ iff there exists a constant L such that

$$\lim_{h\to 0} \frac{|f(x+h) - f(x) - Lh|}{|h|} = 0.$$

The reader should be familiar with the derivative of a function at a point. Differentiability is a *pointwise property* of functions; that is, it is possible for a function to be differentiable at one point and nowhere else (see Example 5.1.5 below).

THEOREM 5.1.4. *Suppose $f : (a, b) \longrightarrow \mathbb{R}$ is differentiable at a point $x \in (a, b)$. Then f is continuous at x.*

Proof. Take $\varepsilon = 1$. Then there exists a $\delta > 0$ such that $|f(x+h) - f(x) - f'(x)h| < \varepsilon|h| = |h|$ whenever $|h| < \delta$. It follows from the triangle inequality that

$$|f(x+h) - f(x)| < |h| + |f'(x)||h| = (1 + |f'(x)|)|h|$$

when $|h| < \delta$. Letting $h \to 0$, we get the result.

EXAMPLE 5.1.5. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x^2, & x \text{ rational}, \\ 0, & x \text{ irrational}. \end{cases}$$

This function is discontinuous at every nonzero x. On the other hand, f is continuous at $x = 0$, and

$$f'(0) = \lim_{h \to 0} \frac{f(0+h) - f(0)}{h} = 0.$$

Hence f is differentiable at $x = 0$ and nowhere else.

EXERCISE 5.1.6. Generalize the function from Exercise 2.4.17. Let $r \geq 1$ and set

$$f_r(x) = \begin{cases} \frac{1}{q^r} & \text{if } x = \frac{p}{q} \text{ in lowest terms and } x \neq 0, \\ 0 & \text{if } x = 0 \text{ or } x \text{ is irrational.} \end{cases}$$

(i) Show that for any $r \geq 1$, f_r is continuous at 0 and at the irrational numbers and is discontinuous at the nonzero rationals.
(ii) If $1 \leq r \leq 2$, show that f_r is not differentiable at any irrational point. (Hint: Use Theorem 1.3.9.)
(iii) For which r is f_r differentiable at $x = 0$?

To settle all discussions about the relationship between differentiability and continuity, consider the following example.

EXAMPLE 5.1.7. We want to create a continuous function on the interval $[0, \infty)$ that is not differentiable at any point in the interior of that interval. Define

$$f_1(x) = \begin{cases} x & \text{if } x \leq 1/2, \\ 1 - x & \text{if } 1/2 \leq x \leq 1 \end{cases}$$

and extend periodically to $[0, \infty)$ by $f_1(x + 1) = f_1(x)$. Then define for all $n \geq 2$, $f_n(x) = \frac{1}{2} f_{n-1}(2x)$. Let $S_m(x) = \sum_{n=1}^{m} f_n(x)$. Then S_m is a continuous function on $[0, \infty)$ and the sequence $(S_m)_{m \in \mathbb{N}}$ converges uniformly to a continuous function S.

EXERCISE 5.1.8. Show that S is not differentiable at any point in $(0, \infty)$.

THEOREM 5.1.9. *Suppose $f : (a, b) \longrightarrow \mathbb{R}$ and $g : (a, b) \longrightarrow \mathbb{R}$ are both differentiable at $x \in (a, b)$. Then for any $\alpha \in \mathbb{R}$, $\alpha f + g$ is differentiable at x. Also, the product fg and the quotient $\frac{f}{g}$ are differentiable at x (for $\frac{f}{g}$ we must have $g(x) \neq 0$). We have:*

(i) $(\alpha f + g)'(x) = \alpha f'(x) + g'(x)$;
(ii) $(fg)'(x) = f(x)g'(x) + f'(x)g(x)$;
(iii) $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$.

Proof. Look in a rigorous calculus book.

Before proceeding further, we want to create a setting which will prevail throughout the theory of differentiation. We said above that differentiability is a pointwise property. Generally speaking, we will assume that a function is not only differentiable at a point, but at all points in a neighborhood of a given point. It is rarely the case that we deal with functions which are differentiable only at a single point.

We want to pay special attention to the procedure for differentiating the composition of functions, sometimes known as the *chain rule*. The proof here takes a little care and will require even more care in several variables.

THEOREM 5.1.10 **(Chain rule)**. *Let f be differentiable at a point a and let g be differentiable at $f(a)$. Then $g \circ f$ is differentiable at a and $D(g \circ f)(a) = (Dg)(f(a))Df(a)$.*

Proof. Let $b = f(a)$, $L = Df(a)$, and $M = Dg(b)$. Set

$$F(x) = f(x) - f(a) - L(x-a),$$
$$G(y) = g(y) - g(b) - M(y-b),$$
$$H(x) = (g \circ f)(x) - (g \circ f)(a) - ML(x-a).$$

By hypothesis, we know that

$$\lim_{x \to a} \frac{|F(x)|}{|x-a|} = \lim_{y \to b} \frac{|G(y)|}{|y-b|} = 0.$$

To prove the theorem, we must show that

$$\lim_{x \to a} \frac{|H(x)|}{|x-a|} = 0.$$

Notice that $H(x) = G(f(x)) + MF(x)$. Now,

$$\frac{|MF(x)|}{|x-a|} = |M| \frac{|F(x)|}{|x-a|} \to 0$$

as $x \to a$. For the remaining term, it follows from above that, given $\varepsilon > 0$, there exists a $\delta > 0$, such that $|y-b| < \delta$ implies $|G(y)| < \varepsilon |y-b|$. The continuity of f at a implies that there exists a $\delta' > 0$ such that $|f(x)-b| < \delta$ when $|x-a| < \delta'$. Hence, if $|x-a| < \delta'$, we have $|G(f(x))| < \varepsilon |f(x)-b|$. But, $|f(x)-b| \leq |F(x)| + |L||x-a|$ so

$$\frac{|G(f(x))|}{|x-a|} \to 0$$

as $x \to a$.

EXERCISE 5.1.11. Give a critique of the following supposed proof of the chain rule:

$$\begin{aligned} \lim_{x \to a} \frac{g(f(x)) - g(f(a))}{x-a} &= \lim_{x \to a} \frac{g(f(x)) - g(f(a))}{f(x) - f(a)} \frac{f(x) - f(a)}{x-a} \\ &= \left(\lim_{x \to a} \frac{g(f(x)) - g(f(a))}{f(x) - f(a)} \right) \left(\lim_{x \to a} \frac{f(x) - f(a)}{x-a} \right) \\ &= Dg(f(a))Df(a). \end{aligned}$$

Suppose that $[a,b]$ is a closed interval in $\mathbb{R}$ and $f : [a,b] \longrightarrow \mathbb{R}$ is continuous. Assume f is differentiable on (a,b). Since f is continuous, we know that f assumes a maximum and a minimum value on $[a,b]$. This observation leads to the following familiar fact from elementary calculus.

THEOREM 5.1.12. *Suppose that f satisfies the hypotheses above and that f assumes a local maximum or a minimum at a point $c \in (a,b)$. Then $f'(c) = 0$.*

Proof. Assume that f has a local maximum at c. There exists $\varepsilon > 0$ such that if $|x - c| < \varepsilon$, then $f(x) \leq f(c)$. It follows that if $x \in (c-\varepsilon, c)$, then $\frac{f(x)-f(c)}{x-c} \geq 0$ and if $x \in (c, c+\varepsilon)$, then $\frac{f(x)-f(c)}{x-c} \leq 0$. The result now follows. The proof for a local minimum is similar.

We now take up various versions of the Mean Value Theorem. The Mean Value Theorem can be regarded as the most important theorem in analysis both in one and several variables. The statements in the following theorem are often called *Rolle's theorem*, the *Mean Value Theorem* (MVT), and the *Generalized Mean Value Theorem*. We state them all here so that the reader will be in familiar territory. We prove only the Generalized Mean Value Theorem, which immediately implies the other two. The Generalized Mean Value Theorem is often called *Cauchy's Mean Value Theorem*.

THEOREM 5.1.13 **(Mean Value Theorem)**. *Let $f : [a,b] \longrightarrow \mathbb{R}$ be continuous and suppose that f is differentiable on (a,b).*

(i) *If $f(a) = f(b)$, then there exists $c \in (a,b)$ such that $f'(c) = 0$.*
(ii) *In any case, there exists $c \in (a,b)$ such that $f'(c)(b-a) = f(b) - f(a)$.*
(iii) *If g satisfies the same hypotheses as f, then there exists $c \in (a,b)$ such that*

$$(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c).$$

Proof. To prove (iii), set $h(x) = (f(b)-f(a))g(x)-(g(b)-g(a))f(x)$. Then h is continuous on $[a,b]$ and differentiable on (a,b). Note that $h(a) = h(b)$. If h is constant, we are done. If not, h assumes a maximum or a minimum at some point $c \in (a,b)$. Theorem 5.1.12 says that $h'(c) = 0$.

The Mean Value Theorem has some serious applications. The first is important for the Fundamental Theorem of Calculus.

COROLLARY 5.1.14. *Suppose f is continuous on $[a,b]$ and differentiable on (a,b). If $f'(x) = 0$ for all $x \in (a,b)$, then f is constant.*

Proof. Given any two points x and y in $[a,b]$ with $x < y$, there exists a point $c \in (x,y)$ such that $f(y)-f(x) = (y-x)f'(c) = 0$. Hence, $f(x) = f(y)$.

COROLLARY 5.1.15. *Suppose f is continuous on $[a,b]$ and differentiable on (a,b). If $f'(c) > 0$ for all $c \in (a,b)$, then f is strictly monotonic increasing.*

Proof. For $x, y \in [a,b]$ with $x < y$, there exists $c \in (x,y)$ so that $f(y) - f(x) = (y-x)f'(c) > 0$.

EXERCISE 5.1.16. Suppose f is continuous on $[a,b]$ and differentiable on (a,b), and suppose $f'(c) < 0$ for all $c \in (a,b)$. Show that f is strictly monotonic decreasing.

EXERCISE 5.1.17 **(L'Hôpital's rule)**. Let (a, b) be any open interval in $\mathbb{R}$, and suppose f and g are differentiable on (a, b). Take $c \in (a, b)$, and suppose $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$ and $\lim_{x\to c} \frac{f'(x)}{g'(x)}$ exists. Show that $\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f'(x)}{g'(x)}$.

We now pause to present one of the all-time favorite examples in one variable differentiation.

EXAMPLE 5.1.18. Let

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{when } x \neq 0, \\ 0 & \text{when } x = 0. \end{cases}$$

Then

$$f'(x) = \begin{cases} 2x \sin(1/x) - \cos(1/x) & \text{when } x \neq 0, \\ 0 & \text{when } x = 0. \end{cases}$$

So even though $f'(0)$ exists, f' is not continuous at 0.

Things are not really as bad as they seem, because, although the derivative may not be continuous, it does have the intermediate value property.

THEOREM 5.1.19 **(Intermediate Value Theorem for Derivatives)**. *Let $f : [a, b] \longrightarrow \mathbb{R}$ be continuous and let f be differentiable on (a, b). Suppose that $(c, d) \subseteq (a, b)$, $f'(c) < 0$, and $f'(d) > 0$. Then there exists a point $x \in (c, d)$ such that $f'(x) = 0$.*

Proof. Since f is continuous on $[c, d]$, it assumes both a maximum and minimum value. Since $f'(c) < 0$, there exists a point $x \in (c, d)$ such that $f(x) < f(c)$ and, since $f'(d) > 0$, there exists a point $y \in (c, d)$ such that $f(y) < f(d)$. Hence the minimum does not occur at either c or d. The conclusion follows.

EXERCISE 5.1.20. There is a point in the previous proof which requires attention. It is related to Corollary 5.1.15 and the exercise which follows it. In the theorem above, it is assumed that $f'(c) < 0$. One might be inclined to think that this means that f is decreasing in a neighborhood of c. To show that this is not true, consider the function

$$f(x) = \begin{cases} 2x^2 \sin(1/x) + x, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

Show that f has a positive derivative at $x = 0$ but is not increasing in any neighborhood of $x = 0$.

There is no reason to stop at one derivative. Once we get started, we can continue taking derivatives as long as the function allows us. Most of the functions encountered in elementary calculus such as polynomials, rational functions, trigonometric functions, exponential functions, logarithmic functions, hyperbolic functions, and algebraic functions, are differentiable

infinitely often as long as nothing untoward happens in the domain. The above functions make up more or less the entire list of functions considered in elementary calculus. We assume that the reader knows how to differentiate them. The following definition is useful throughout analysis.

DEFINITION 5.1.21. Let f be a continuous function from (a,b) to $\mathbb{R}$. If k is an integer greater than or equal to 1, we say that $f \in C^k(a,b)$ if f has k derivatives at each point in (a,b) and each of these derivatives is continuous on (a,b). Elements of $C^1(a,b)$ are said to be *continuously differentiable* on (a,b).

We denote the k-th derivative of f by $f^{(k)}$. By convention, we take $f^{(0)} = f$. We say that $f \in C^\infty(a,b)$ if f has derivatives of all orders on (a,b). If U is any open set in $\mathbb{R}$, the expressions $C^k(U)$ and $C^\infty(U)$ are defined similarly.

EXERCISE 5.1.22. Suppose $f \in C^1(a,b)$. Let $[c,d] \subseteq (a,b)$. Then there exists a constant M such that for all $x, y \in [c,d]$, we have $|f(y) - f(x)| \leq M|y - x|$.

EXERCISE 5.1.23. Let

$$f(x) = \begin{cases} e^{-1/x^2} & \text{when } x \neq 0, \\ 0 & \text{when } x = 0. \end{cases}$$

(i) Show that $f \in C^\infty(\mathbb{R})$.
(ii) Using L'Hôpital's rule, or anything you wish, show that $f^{(k)}(0) = 0$ for all $k \geq 0$.

COROLLARY 5.1.24 **(Taylor's theorem)**. *Suppose $f \in C^{k+1}(a,b)$ and $x_0 \in (a,b)$. Then, for any $x \in (a,b)$, we can write*

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 \\ &\quad + \cdots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(k+1)}(c)}{(k+1)!}(x - x_0)^{k+1} \end{aligned}$$

where c is some point between x and x_0.

Proof. Without loss of generality, we assume that $x > x_0$. For $t \in [x_0, x]$, define a polynomial $T_k(x,t)$ in the variable x by

$$T_k(x,t) = f(t) + f'(t)(x - t) + \frac{f''(t)}{2!}(x - t)^2 + \cdots + \frac{f^{(k)}(t)}{k!}(x - t)^k.$$

Define $R_k(x,t) = f(x) - T_k(x,t)$. Note that we have $R_k(x, x_0) \in C^{k+1}(x_0, b)$, $R_k^{(i)}(x_0, x_0) = 0$ for $0 \leq i \leq k$, and $R_k^{(k+1)}(x, x_0) = f^{(k+1)}(x)$ for all $x \in (a,b)$. For $x > x_0$ and $t \in [x_0, x]$, set

$$Q(t) = (x - t)^{k+1} \frac{R_k(x, x_0)}{(x - x_0)^{k+1}} - R_k(x,t).$$

It is clear that Q is continuous on $[x_0, x]$, differentiable on (x_0, x), and $Q(x_0) = Q(x) = 0$. It is easy to see that

$$Q'(t) = -(k+1)(x-t)^k \frac{R_k(x, x_0)}{(x-x_0)^{k+1}} + \frac{(x-t)^k}{k!} f^{(k+1)}(t).$$

Hence, by Rolle's theorem, there exists $c \in (x_0, x)$ such that

$$(k+1)(x-c)^k \frac{R_k(x, x_0)}{(x-x_0)^{k+1}} = \frac{(x-c)^k}{k!} f^{(k+1)}(c).$$

What Taylor's theorem allows us to do is to approximate a C^{k+1} function in the neighborhood of a point by a polynomial of degree k. Usually, the *remainder term*,

$$R_k(x, x_0) = \frac{f^{(k+1)}(c)}{(k+1)!}(x-x_0)^{k+1},$$

is reasonably small because of the $(k+1)!$ in the denominator. The expansion in Taylor's theorem is called the *Taylor expansion of f about x_0*.

EXERCISE 5.1.25. Find the Taylor expansions of the following functions about the indicated points to at least six terms:

(i) $f(x) = \sin(x)$ about $x_0 = \pi$;
(ii) $f(x) = \frac{1}{x-1}$ about $x_0 = -1$;
(iii) $f(x) = e^{-1/x^2}$ about $x_0 = 0$;
(iv) $f(x) = \sqrt{x^2+1}$ about $x_0 = 2$.

EXERCISE 5.1.26.

(i) Suppose that $f \in C^1(a, b)$ and $f'(x_0) = 0$ for some $x_0 \in (a, b)$. Show that f may or may not have a local extremum at x_0.
(ii) Suppose that $f \in C^2(a, b)$, $f'(x_0) = 0$, and $f''(x_0) > 0$. Show that f has a local minimum at x_0. Formulate and prove an analogous statement in case $f''(x_0) < 0$.
(iii) Suppose that $f \in C^{k+1}(a, b)$, $f^{(i)}(x_0) = 0$ for $1 \le i \le k$, and $f^{(k+1)}(x_0) \neq 0$. Under what conditions can you say that f must have a local extremum at x_0?

As another application of the Mean Value Theorem, we present the theorem of Liouville. This theorem can sometimes be used to determine whether a given irrational real number is algebraic or transcendental.

DEFINITION 5.1.27. If α is a real algebraic number, then we say that α *is algebraic of degree n* if n is the smallest degree of a nonzero polynomial in $\mathbb{Z}[x]$ of which α is a root.

THEOREM 5.1.28 **(Liouville's theorem)**. *Let $\alpha \in \mathbb{R}$ be algebraic of degree $n \ge 2$. Then there exists $C = C(\alpha)$ depending on α such that $|\alpha - p/q| > C/q^n$ for all $p/q \in \mathbb{Q}$.*

Proof. Let $f(x) \in \mathbb{Z}[x]$ be an irreducible polynomial of degree n such that $f(\alpha) = 0$. Let $p/q \in \mathbb{Q}$ and assume that $\alpha < p/q$. By the Mean Value Theorem, there exists c with $\alpha < c < p/q$ such that $f(p/q) - f(\alpha) = f(p/q) = (p/q - \alpha)f'(c)$. By the irreducibility of $f(x)$, we have $f(p/q) \neq 0$, and therefore $f'(c) \neq 0$. Assume that $0 < p/q - \alpha < 1$. Choose $d > 0$ such that $|f'(x)| < 1/d$ for $x \in [\alpha, p/q]$. Then $|f(p/q)| = |p/q - \alpha|\,|f'(c)| \neq 0$. Since $f(x)$ has integer coefficients, $|q^n f(p/q)| \in \mathbb{Z}$, so it is greater than or equal to 1. Hence, $1 \leq q^n|p/q - \alpha|(1/d)$. It follows that $|\alpha - p/q| \geq d/q^n$ when $0 < p/q - \alpha < 1$.

Exercise 5.1.29.

(i) Modify the above proof to cover the case when $p/q < \alpha$.

(ii) Show that there exists a constant C such that $|\alpha - p/q| > C/q^n$ for all $p/q \in \mathbb{Q}$.

Exercise 5.1.30. Suppose $n \geq 2$ and α is real algebraic of degree n. Show that if $r > n$, then the function f_r of Exercise 5.1.6 is differentiable at α.

In elementary calculus, the derivative is often motivated through a discussion of the tangent line to a curve at a point. This is accomplished using secant lines which approximate the tangent line and then taking limits. The actual definition of the tangent line to the graph of a function at a point is as follows: if a function is differentiable at a point, the tangent line at that point is the line through the point whose slope is the derivative of the function at that point. We can think of the tangent line as the *best linear approximation* at that point. This is the idea that motivates the concept of derivative in several variables.

5.2. Differential Calculus in $\mathbb{R}^n$

We now turn to a study of the properties of differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. The definition of a differentiable function is motivated by the discussion at the end of the previous section about the derivative of a function $f : \mathbb{R} \longrightarrow \mathbb{R}$. As in the case of the derivative of functions of one variable, we form a difference quotient, which means that we must divide by something. Since division in $\mathbb{R}^n$ does not make sense for $n > 1$, we keep the divisor in $\mathbb{R}$. Throughout this chapter, we use the Euclidean norm in $\mathbb{R}^n$. We also recall that if $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a linear map, then T is bounded.

Definition 5.2.1. Suppose $U \subseteq \mathbb{R}^n$ is an open set. A function $f : U \longrightarrow \mathbb{R}^m$ is *differentiable at* $x \in U$ if there is a linear map $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - Th\|}{\|h\|} = 0.$$

First, note that $h \to 0$ in $\mathbb{R}^n$. Notice also that the norm sign in the numerator denotes the Euclidean norm in $\mathbb{R}^m$ while the norm sign in the denominator

denotes the Euclidean norm in $\mathbb{R}^n$. If we use the norm sign for an element of $\mathbb{R}$, it indicates the usual absolute value on $\mathbb{R}$. We write $T = Df(x)$ and we call this the *derivative* of f at x. We say that f is *differentiable on* U if f is differentiable at each point in U.

Thus the derivative of a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ at a point is a linear transformation. It may be difficult to think of this as a generalization of the slope of a tangent line as it is in one variable. However, if one thinks of the tangent line in one variable as the best linear approximation to a function at a point, we can think of the derivative in $\mathbb{R}^n$ as a generalization of this concept; that is,

$$f(x_0) + Df(x_0)(x - x_0)$$

provides the "best linear approximation" to the function f at the point $x_0 \in \mathbb{R}^n$.

There are many theorems about derivatives of functions of several variables which are analogous to those in one variable.

THEOREM 5.2.2. *Suppose U is an open set in $\mathbb{R}^n$ and $f : U \longrightarrow \mathbb{R}^m$ is differentiable at a point $x_0 \in U$. Then f is continuous at x_0.*

Proof. Take $\varepsilon = 1$. Then there exists a $\delta > 0$ such that

$$\|f(x_0 + h) - f(x_0) - Df(x_0)h\| < \varepsilon\|h\| = \|h\|$$

whenever $\|h\| < \delta$. It follows from the triangle inequality that

$$\begin{aligned}\|f(x_0 + h) - f(x_0)\| &< \|h\| + \|Df(x_0)h\| \\ &\leq \|h\| + \|Df(x_0)\| \cdot \|h\| = (1 + \|Df(x_0)\|)\|h\|\end{aligned}$$

when $\|h\| < \delta$.

When $m = 1$ and f is a real-valued function, we are led to a special situation.

DEFINITION 5.2.3. Let $U \subseteq \mathbb{R}^n$ be an open set, and let $f : U \longrightarrow \mathbb{R}$ be differentiable on U. Let $x \in U$ and let $v \in \mathbb{R}^n$ be a unit vector. The *directional derivative* of f at x in the direction v is defined as

$$D_v f(x) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t}.$$

In the particular case when $v = e_j$, a standard basis vector, we obtain the *partial derivative in the j-th direction*

$$D_j f(x) = D_{e_j} f(x) = \lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t}.$$

EXERCISE 5.2.4.

(i) Let $U \subseteq \mathbb{R}^n$ be an open set and let $f : U \longrightarrow \mathbb{R}^m$ be differentiable. Write $f = (f_1, f_2, \ldots, f_m)$, where $f_k : U \longrightarrow \mathbb{R}$ is the k-th coordinate function of f. Let $v \in \mathbb{R}^n$ be a unit vector. If $x \in U$, define

$$D_v f(x) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t}.$$

Show that $D_v f(x)$ exists if and only if $D_v f_k(x)$ exists for all k, $1 \leq k \leq m$, and, in this case, $D_v f(x) = (D_v f_1(x), D_v f_2(x), \ldots, D_v f_m(x))$.

(ii) Explain why it is useful for us to have required v to be a unit vector.

REMARK 5.2.5. Note that, in Definition 5.2.3, these directional derivatives are real honest-to-goodness derivatives of functions of one variable, that is, they represent the rate of change of a function in a particular direction. The partial derivatives play a special role, as we shall see below. It is worth observing that the classical notation for $D_j f(x)$ is $\frac{\partial f}{\partial x_j}(x)$. All sorts of theorems and properties can be stated much more easily with the notation $D_j f(x)$.

EXERCISE 5.2.6.

(i) Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be defined by $f(x_1, x_2, \ldots, x_n) = x_k$ for some k, for $1 \leq k \leq n$. Show that f is differentiable at any point and that $Df(x) = f$ for all $x \in \mathbb{R}^n$.

(ii) Find $D_v f(x)$ for any unit vector v in $\mathbb{R}^n$.

REMARK 5.2.7. The map f in the above exercise is called the *projection onto the k-th coordinate* and is denoted p_k. More generally, if $m \leq n$, we can pick indices $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ and define a projection $p : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ by $p(x_1, x_2, \ldots, x_n) = (x_{i_1}, x_{i_2}, \ldots, x_{i_m})$.

EXERCISE 5.2.8. Show that any such p as above is differentiable and find its derivative.

All of the statements below are easy exercises, but we prove one or two just to show that we are working at this.

PROPOSITION 5.2.9. *If U is an open set in $\mathbb{R}^n$ and $f : U \longrightarrow \mathbb{R}^m$ is differentiable, then the derivative of f is unique.*

Proof. Suppose T and T' are derivatives of f at $x \in U$. Then, for $h \in \mathbb{R}^n$,

$$\begin{aligned} \frac{\|Th - T'h\|}{\|h\|} &= \frac{\|(f(x+h) - f(x) - Th) - (f(x+h) - f(x) - T'h)\|}{\|h\|} \\ &\leq \frac{\|f(x+h) - f(x) - Th\|}{\|h\|} + \frac{\|f(x+h) - f(x) - T'h\|}{\|h\|}. \end{aligned}$$

Thus $\|Th - T'h\|/\|h\| \to 0$ as $h \to 0$. But $T - T'$ is a linear map, so by the definition of the norm of an operator,

$$\max_{\|h\|=M} \frac{\|Th - T'h\|}{\|h\|} = \|T - T'\|.$$

Because this is a constant independent of M for $M \neq 0$ we must have $\|T - T'\| = 0$, that is, $T = T'$.

EXERCISE 5.2.10. If $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is a linear map and $x \in \mathbb{R}^n$, then f is differentiable at x and $Df(x) = f$.

This exercise says something which is almost tautological, but it says something nonetheless. That is, if $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is linear and $h \in \mathbb{R}^n$, then at any point $x \in \mathbb{R}^n$, $Df(x)h = f(h)$. How does this all work for a linear function $f : \mathbb{R} \longrightarrow \mathbb{R}$? Notice first that if $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a linear map, then $f(0)$ must be equal to 0. Moreover, there is an element $a \in \mathbb{R}$ such that $f(x) = ax$, for all $x \in \mathbb{R}$. Conversely, given an element $a \in \mathbb{R}$ we can construct a linear map $f_a : \mathbb{R} \longrightarrow \mathbb{R}$ defined by $f_a(x) = ax$. In elementary calculus, we use this correspondence to treat derivatives of functions of one variable as numbers at each point, or as functions on $\mathbb{R}$, rather than as linear maps $\mathbb{R} \longrightarrow \mathbb{R}$ at each point. In our present case, instead of saying that if $f_a(x) = ax$, then $f_a'(x) = a$, we say that $Df_a(x) = f_a$ for all x. Observe that this discussion tells us that if a function is already linear, then the best linear approximation is the function itself.

More generally, if $f : \mathbb{R} \longrightarrow \mathbb{R}^n$ is a function, we often think of the derivative of f at a point x simply as a vector in $\mathbb{R}^n$. By analogy with the real-valued situation, we use the notation $f'(x)$ and define

$$f'(x) = Df(x)(1).$$

PROPOSITION 5.2.11. *Let U be an open set in $\mathbb{R}^n$ and let $f, g : U \longrightarrow \mathbb{R}^m$ be differentiable on U. Then $f+g$ is differentiable on U, and $D(f+g)(x) = Df(x) + Dg(x)$ for $x \in U$.*

Proof. For $x \in U$,

$$\begin{aligned} &\frac{\|(f+g)(x+h) - (f+g)(x) - (Df(x) + Dg(x))h\|}{\|h\|} \\ &\leq \frac{\|f(x+h) - f(x) - Df(x)h\|}{\|h\|} + \frac{\|g(x+h) - g(x) - Dg(x)h\|}{\|h\|}. \end{aligned}$$

The reader might ask about the product rule. The question depends on the type of multiplication being used. Let us try the easiest case in which f and g map U to $\mathbb{R}$.

PROPOSITION 5.2.12. *If U is an open set in $\mathbb{R}^n$ and if $f : U \longrightarrow \mathbb{R}$ and $g : U \longrightarrow \mathbb{R}$ are real-valued functions which are differentiable on U, then fg is differentiable on U. For $x \in U$, we have $D(fg)(x) = f(x)Dg(x) + g(x)Df(x)$.*

Proof. Before starting the proof, we observe that $f(x)Dg(x)$ really makes sense, since $f(x)$ is a real scalar and $Dg(x)$ is a linear map, and we can always multiply linear maps by scalars. So let's go:

$$\begin{aligned}
&\frac{1}{\|h\|}\|f(x+h)g(x+h) - f(x)g(x) - (f(x)Dg(x) + g(x)Df(x))h\| \\
&\quad = \frac{1}{\|h\|}\|(f(x+h)[g(x+h) - g(x)] - f(x)Dg(x)h) \\
&\qquad\qquad + (g(x)[f(x+h) - f(x)] - g(x)Df(x)h)\| \\
&\quad \leq \frac{1}{\|h\|}\big(\|(f(x+h) - f(x))(g(x+h) - g(x))\| \\
&\qquad\qquad + \|f(x)(g(x+h) - g(x) - Dg(x)h)\| \\
&\qquad\qquad + \ \|g(x)(f(x+h) - f(x) - Df(x)h)\|\big).
\end{aligned}$$

By the definition of the derivative, the second and third summands in the last expression vanish as $h \to 0$. For the first term, we have

$$\begin{aligned}
&\frac{1}{\|h\|}\|(f(x+h) - f(x))(g(x+h) - g(x))\| \\
&\qquad = \frac{\|f(x+h) - f(x)\|}{\|h\|}\|g(x+h) - g(x)\|.
\end{aligned}$$

EXERCISE 5.2.13. Finish the proof by showing $\frac{\|f(x+h)-f(x)\|}{\|h\|}\|g(x+h) - g(x)\|$ goes to zero as $h \to 0$.

We now turn to the chain rule, which is a very important theorem in the calculus of several variables. The reader will note that the proof is pretty much the same as the proof in one variable (Theorem 5.1.10).

THEOREM 5.2.14 **(Chain rule)**. *Suppose that U is an open set in $\mathbb{R}^n$ and $f : U \longrightarrow \mathbb{R}^m$ is differentiable on U. Let V be an open set in $\mathbb{R}^m$ such that $f(U) \subseteq V$. Suppose $g : V \longrightarrow \mathbb{R}^p$ is differentiable on V. Then $g \circ f : U \longrightarrow \mathbb{R}^p$ is differentiable on U, and, for any $a \in U$, we have $D(g \circ f)(a) = Dg(f(a)) \circ Df(a)$.*

Proof. Let $b = f(a)$, $L = Df(a)$, and $M = Dg(b)$. Set

$$\begin{aligned}
\phi(x) &= f(x) - f(a) - L(x - a), \\
\psi(y) &= g(y) - g(b) - M(y - b),
\end{aligned}$$

and

$$\rho(x) = (g \circ f)(x) - (g \circ f)(a) - ML(x - a).$$

Then, our hypotheses tell us that

$$\lim_{x\to a}\frac{\|\phi(x)\|}{\|x-a\|}=\lim_{y\to b}\frac{\|\psi(y)\|}{\|y-b\|}=0,$$

and we need to show that

$$\lim_{x\to a}\frac{\|\rho(x)\|}{\|x-a\|}=0.$$

But

$$\rho(x)=\psi(f(x))+M\phi(x)$$

(write out the definitions). Since M is a bounded linear map,

$$\frac{\|M\phi(x)\|}{\|x-a\|}\le\|M\|\frac{\|\phi(x)\|}{\|x-a\|}\to 0$$

as $x\to a$. By the above limit and the continuity of f, for any $\varepsilon>0$, there is $\delta>0$ such that $\|x-a\|<\delta$ ensures $\|\psi(f(x))\|<\varepsilon\|f(x)-b\|$. But $\|f(x)-b\|\le\|\phi(x)\|+\|L\|\|x-a\|$, so

$$\frac{\|\psi(f(x))\|}{\|x-a\|}\to 0$$

as $x\to a$ as well.

5.3. The Derivative as a Matrix of Partial Derivatives

We now proceed to some of the more computational aspects of differentiation theory in several variables. Let us consider the form of a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. If $f:\mathbb{R}^n\longrightarrow\mathbb{R}^m$, we can write $f(x)=(f_1(x),f_2(x),\ldots,f_m(x))$. Then, for each k with $1\le k\le m$, f_k is a function $\mathbb{R}^n\longrightarrow\mathbb{R}$. These functions f_k are the *component functions* of f, and there is an obvious fact about them.

COROLLARY 5.3.1 **(To the chain rule)**. *If f is differentiable, then f_k is differentiable for each k.*

Proof. Observe that $f_k=p_k\circ f$ and use the chain rule.

EXERCISE 5.3.2. Show that if each f_k is differentiable on an open set $U\subseteq\mathbb{R}^n$, then $f=(f_1,f_2,\ldots,f_m)$ is differentiable on U and $Df=(Df_1,Df_2,\ldots,Df_m)$.

How are we interpreting $(Df_1,Df_2,\ldots,Df_m)$ as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$? For $x\in U$ the expression

$$(Df_1(x),Df_2(x),\ldots,Df_m(x))$$

is the linear transformation whose value at $y\in\mathbb{R}^n$ is given by

$$[(Df_1(x),Df_2(x),\ldots,Df_m(x))](y)=(Df_1(x)(y),Df_2(x)(y),\ldots,Df_m(x)(y)).$$

What do partial derivatives have to do with all this? From Appendix B, we know that linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be represented by

matrices with respect to any bases we choose in the respective spaces. Let us work first with a differentiable function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. We know that

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - Df(x)h\|}{\|h\|} = 0.$$

If we think of $Df(x)$ as a $1 \times n$ matrix of the form $(a_1(x), a_2(x), \ldots, a_n(x))$ and write $h = (h_1, h_2, \ldots, h_n)$, then

$$Df(x)h = (a_1(x), a_2(x), \ldots, a_n(x)) \cdot (h_1, h_2, \ldots, h_n),$$

where the dot signifies the dot product in $\mathbb{R}^n$. Now consider the case $h = te_j$. It follows from Definition 5.2.3 that $a_j(x) = D_j f(x)$. This is summarized in the following statement.

PROPOSITION 5.3.3. *If $U \subseteq \mathbb{R}^n$ is open and $f : U \longrightarrow \mathbb{R}$ is differentiable on U, then*

$$Df(x) = (D_1 f(x), D_2 f(x), \ldots, D_n f(x)).$$

For $h \in \mathbb{R}^n$,

$$Df(x)h = D_1 f(x)h_1 + D_2 f(x)h_2 + \cdots + D_n f(x)h_n.$$

EXERCISE 5.3.4. Let $U \subseteq \mathbb{R}^n$ be open, and let $f = (f_1, f_2, \ldots, f_m) : U \longrightarrow \mathbb{R}^m$ be differentiable on U. If $x \in U$ and (a_{ij}) is the matrix of $Df(x)$ with respect to the standard basis, then $a_{ij} = D_j f_i(x)$.

DEFINITION 5.3.5. Let $U \subseteq \mathbb{R}^n$ be open, let $f : U \longrightarrow \mathbb{R}^m$ be differentiable on U, and let $x \in U$. The matrix $(D_j f_i(x))_{1\le i\le m, 1\le j\le n}$ is called the *Jacobian matrix* of f at x.

In case $m = 1$, the vector-valued function $(D_1 f(x), D_2 f(x), \ldots, D_n f(x))$ is called the *gradient vector* and is denoted $\nabla f(x)$. Thus, by Proposition 5.3.3, $Df(x)(h) = \nabla f(x) \cdot h$.

EXERCISE 5.3.6.

(i) Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \longrightarrow \mathbb{R}$ be differentiable on U. If $x \in U$ and v is a unit vector in $\mathbb{R}^n$, then $D_v f(x) = \nabla f(x) \cdot v$.

(ii) Show that the maximum value of the directional derivative of f at a point x is in the direction of $\nabla f(x)$ and the value of this directional derivative is $\|\nabla f(x)\|$.

The existence of the partial derivatives at a point, or even in the neighborhood of a point, does not assure the differentiability of the function at the point. Actually, it is much worse than that. As shown in the following example, the existence of the directional derivative at a point in every direction does not assure that the function is differentiable at the point. For differentiability, one must approach the given point from every direction along all sorts of paths, while the directional derivative is taken only along straight-line paths through the point. See Challenge Problem 6.1.

EXAMPLE 5.3.7. Let

$$f(x,y) = \begin{cases} 0 & \text{when } (x,y) = (0,0), \\ \frac{xy}{x^2+y^2} & \text{otherwise.} \end{cases}$$

Then $D_1 f(0,0) = D_2 f(0,0) = 0$, but nonetheless f is not continuous at the origin, and hence not differentiable at $(0,0)$.

EXERCISE 5.3.8.

(i) Let

$$f(x,y) = \begin{cases} 0 & \text{when } (x,y) = (0,0), \\ \frac{x^3}{x^2+y^2} & \text{otherwise.} \end{cases}$$

Show that f has a directional derivative in every direction at the origin but that f is not continuous at the origin, and hence not differentiable.

(ii) Let

$$f(x,y) = \begin{cases} 0 & \text{when } (x,y) = (0,0), \\ (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) & \text{otherwise.} \end{cases}$$

Show that $D_1 f$ and $D_2 f$ exist everywhere but are not continuous at the origin.

THEOREM 5.3.9. *Let U be an open set in $\mathbb{R}^n$ and let $f : U \longrightarrow \mathbb{R}^m$ be a function with the property that $D_j f_i$ is continuous on U for $1 \le i \le m$, $1 \le j \le n$. Then f is differentiable on U and, as we might expect, $Df(x) = (D_j f_i(x))_{1 \le i \le m, 1 \le j \le n}$.*

Proof. We prove it first for $n = 2$ and $m = 1$. The proof uses the Mean Value Theorem for functions of one variable. For $x = (x_1, x_2) \in U$ and $h = (h_1, h_2)$, write

$$\begin{aligned}
&\frac{1}{\|h\|}\|f(x+h) - f(x) - (Df(x))h\| \\
&= \frac{1}{\|h\|}\|f(x_1+h_1, x_2+h_2) - f(x_1,x_2) - (D_1 f(x_1,x_2), D_2 f(x_1,x_2))(h_1,h_2)\| \\
&= \frac{1}{\|h\|}\|f(x_1+h_1, x_2+h_2) - f(x_1, x_2+h_2) - D_1 f(x_1,x_2)h_1 \\
&\qquad + f(x_1, x_2+h_2) - f(x_1,x_2) - D_2 f(x_1,x_2)h_2\| \\
&\le \frac{1}{\|h\|}\|f(x_1+h_1, x_2+h_2) - f(x_1, x_2+h_2) - D_1 f(x_1,x_2)h_1\| \\
&\qquad + \frac{1}{\|h\|}\|f(x_1, x_2+h_2) - f(x_1,x_2) - D_2 f(x_1,x_2)h_2\|.
\end{aligned}$$

Now, by the Mean Value Theorem for functions of one variable, there exists $\xi_1 \in (x_1, x_1+h_1)$ (or $\xi_1 \in (x_1+h_1, x_1)$ if $h_1 < 0$) and $\xi_2 \in (x_2, x_2+h_2)$ (or $\xi_2 \in (x_2+h_2, x_2)$ if $h_2 < 0$) such that

$$f(x_1+h_1, x_2+h_2) - f(x_1, x_2+h_2) = D_1 f(\xi_1, x_2+h_2)h_1$$

and

$$f(x_1, x_2 + h_2) - f(x_1, x_2) = D_2 f(x_1, \xi_2) h_2.$$

Thus the above sequence of inequalities continues as

$$\leq \frac{|h_1|}{\|h\|} \|D_1 f(\xi_1, x_2 + h_2) - D_1 f(x_1, x_2)\| + \frac{|h_2|}{\|h\|} \|D_2 f(x_1, \xi_2) - D_2 f(x_1, x_2)\|$$
$$\leq \|D_1 f(\xi_1, x_2 + h_2) - D_1 f(x_1, x_2)\| + \|D_2 f(x_1, \xi_2) - D_2 f(x_1, x_2)\|$$

and this goes to zero as $h \to 0$ since $D_1 f$ and $D_2 f$ are continuous at (x_1, x_2).

The general case for arbitrary n and m is easy to complete by adding and subtracting enough times and using the Mean Value Theorem over and over again.

So far, we have not computed a lot of derivatives because it is awkward to compute the linear transformation associated to the definition. With this last theorem, it becomes much easier to compute the derivative of a function given by explicit formulas. For example, let $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ be defined by

$$f(x, y, z) = (x^2 y + e^{xz}, \sin(xyz)).$$

Then f is differentiable on all of $\mathbb{R}^3$ and

$$Df(x, y, z) = \begin{pmatrix} 2xy + ze^{xz} & x^2 & xe^{xz} \\ yz\cos(xyz) & xz\cos(xyz) & xy\cos(xyz) \end{pmatrix}.$$

Thus,

$$Df(1, 0, -1) = \begin{pmatrix} -e^{-1} & 1 & e^{-1} \\ 0 & -1 & 0 \end{pmatrix},$$

which represents a linear transformation $\mathbb{R}^3 \longrightarrow \mathbb{R}^2$. The more general expression above assigns to each point $(x, y, z) \in \mathbb{R}^3$ a linear transformation $Df(x, y, z) \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$. Thus, we can think of Df as a function

$$Df : \mathbb{R}^3 \longrightarrow \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2).$$

This last example should totally demystify the idea of derivatives of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. There is of course some interesting theory related to differentiation, and we are in the midst of that exposition. However, to compute partial derivatives, and hence derivatives, requires nothing more than being able to differentiate functions of one variable, which you learned in elementary calculus. In fact, as we have commented before, there really are not that many functions that you can differentiate explicitly. So you will notice that all of the examples involve polynomials, rational functions, trigonometric functions, logarithmic functions, etc.

EXERCISE 5.3.10. For any function $f : \mathbb{C} \longrightarrow \mathbb{C}$, we can write f in terms of its real and imaginary parts as $f(z) = u(x, y) + iv(x, y)$, where u and v are functions $\mathbb{R}^2 \longrightarrow \mathbb{R}$, and z is written in the form $z = x + iy$.

A function $f : \mathbb{C} \longrightarrow \mathbb{C}$ (with the usual metric on $\mathbb{C}$) is *complex differentiable* at $z_0 \in \mathbb{C}$ if

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists. A function f is *analytic* on an open set $U \subseteq \mathbb{C}$ if f is differentiable at each point of U.

(i) Suppose f is analytic on an open set $U \subseteq \mathbb{C}$. Show that u and v are differentiable on U considered as a subset of $\mathbb{R}^2$.
(ii) Suppose f is analytic on an open set $U \subseteq \mathbb{C}$. Show that $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ and $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$. These are called the *Cauchy-Riemann equations.*
(iii) If $U \subseteq \mathbb{C}$ is an open set and u and v are continuously differentiable on U and satisfy the Cauchy-Riemann equations, show that $f(z) = u(x, y) + iv(x, y)$ is analytic on U.
(iv) Find an example of a function $f : \mathbb{C} \longrightarrow \mathbb{C}$ that is differentiable at one point but not in a neighborhood of that point.

EXERCISE 5.3.11. Let $f : \mathbb{C} \longrightarrow \mathbb{C}$ be given by $f(z) = e^z$, which can be written $f(x + iy) = e^x \cos y + ie^x \sin y$ (see Definition 1.9.20). Show that f is analytic on $\mathbb{C}$ and that $f'(z) = f(z)$.

EXERCISE 5.3.12. Let $z_0 \in \mathbb{C}$, and define $f : \mathbb{C} \longrightarrow \mathbb{C}$ by $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$, where $a_n \in \mathbb{C}$ for all n. Let r be the radius of convergence of this power series (see the project in Section 1.10.2) and suppose that $r > 0$.

(i) Show that $f(z)$ is analytic on $B_r(z_0) = \{z \in \mathbb{C} \mid |z - z_0| < r\}$. (Hint: Show that the series can be differentiated term by term inside the radius of convergence.)
(ii) Show that the radius of convergence of the power series for $f'(z)$ is equal to r.

EXERCISE 5.3.13. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be defined by $f(x, y) = \sqrt{|x| + |y|}$. Find those points in $\mathbb{R}^2$ at which f is differentiable.

EXERCISE 5.3.14. Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function such that $|f(x)| \leq \|x\|^\alpha$ for some $\alpha > 1$. Show that f is differentiable at 0.

EXERCISE 5.3.15. Let $f : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ be defined by $f(x, y) = x \cdot y$.

(i) Show that f is differentiable on $\mathbb{R}^n \times \mathbb{R}^n$.
(ii) Show that $Df(a, b)(x, y) = a \cdot y + b \cdot x$.

5.4. The Mean Value Theorem

Next we consider the Mean Value Theorem for functions of several variables. As in the case of functions of one variable, the Mean Value Theorem relates the average rate of change of a function in a specified direction to the instantaneous rate of change at a particular point as measured by the derivative.

In the case of a function of one variable, there is little choice about the so-called specified direction. But, when we consider functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, we find that it is necessary to specify a direction both in the domain and in the range of the function in order to give a proper interpretation to the Mean Value Theorem.

THEOREM 5.4.1. *Let U be an open set in $\mathbb{R}^n$, and let $f : U \longrightarrow \mathbb{R}$ be differentiable on U. Let x, y be two distinct points in U such that the line segment joining x to y lies entirely in U. Then there exists $\xi \in (0,1)$ such that*

$$f(y) - f(x) = Df(z)(y - x),$$

where $z = (1 - \xi)x + \xi y$.

Proof. Define

$$F(t) = f((1 - t)x + ty).$$

Then F is continuous on $[0,1]$ and differentiable on $(0,1)$, so there exists $\xi \in (0,1)$ such that

$$F(1) - F(0) = F'(\xi).$$

The left-hand side is $f(y) - f(x)$, and by the chain rule,

$$F'(\xi) = Df((1 - \xi)x + \xi y)(y - x).$$

Note that $Df(z)(y - x) = D_u f(z)\|y - x\|$, where u is the unit vector in the direction of the vector $y - x$.

EXERCISE 5.4.2. Let U be a connected open set in $\mathbb{R}^n$, and let $f : U \longrightarrow \mathbb{R}$ be differentiable on U. If $Df(x) = 0$ for all $x \in U$, then f is constant on U.

We note again that the Mean Value Theorem for real-valued functions of several variables that we have just proved is really a one-variable theorem. That is, to make sense of the mean value property, it was essential that we move away from a point x in exactly one direction, namely, the straight-line direction from x to y. It is this idea that motivates the statement of a Mean Value Theorem for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. To retain the one-variable nature of the Mean Value Theorem, in addition to having a straight-line direction implicitly chosen for us in the domain (namely, the direction $y - x$ in $\mathbb{R}^n$), we must also explicitly choose a direction in $\mathbb{R}^m$ in order to make sense of the mean value property.

THEOREM 5.4.3. *Let U be an open subset in $\mathbb{R}^n$ and let $f : U \longrightarrow \mathbb{R}^m$ be differentiable on U. For any two distinct points $x, y \in U$ such that the line segment joining x to y lies entirely in U and for any vector $v \in \mathbb{R}^m$, there exists $\xi \in (0,1)$ such that*

$$v \cdot (f(y) - f(x)) = v \cdot Df(z)(y - x),$$

where $z = (1 - \xi)x + \xi y$.

EXERCISE 5.4.4. Prove this.

We note in connection to the discussion above that if v is a unit vector, the expression $v \cdot (f(y) - f(x))$ is the component of $f(y) - f(x)$ in the direction of the vector v. A similar statement is true for the expression $v \cdot Df(z)(y - x)$.

EXERCISE 5.4.5. Let U be a connected open set in $\mathbb{R}^n$, and let $f : U \longrightarrow \mathbb{R}^m$ be differentiable on U. If $Df(x) = 0$ for all $x \in U$, then f is constant on U.

The significance of the multivariable version of the Mean Value Theorem is that the direction vector v is arbitrary. However, in general, there is no single ξ that will satisfy the conclusion of the Mean Value Theorem for all v simultaneously.

EXERCISE 5.4.6. Find an example of a differentiable function $f : \mathbb{R} \longrightarrow \mathbb{R}^2$ that does not satisfy the property that, given any $x, y \in \mathbb{R}$ with $x < y$, there exists $z \in (x, y)$ such that $f(y) - f(x) = Df(z)(y - x)$.

EXERCISE 5.4.7. Let U be an open set in $\mathbb{R}^n$, and let $f : U \longrightarrow \mathbb{R}^m$ be differentiable on U with continuous partial derivatives. (In the next section, we call such a function C^1.) Suppose that B is a compact, convex subset of U. Then there exists a constant M such that for any two points $x, y \in B$, we have $\|f(y) - f(x)\| \leq M\|y - x\|$. This is analogous to Exercise 5.1.22 for functions of one variable.

5.5. Higher-Order Partial Derivatives and Taylor's Theorem

The next natural question is, "What about second derivatives?" There are two ways to look at this. If $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is differentiable at a point x, then $Df(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$. That is, it is a linear transformation $\mathbb{R}^n \longrightarrow \mathbb{R}^m$. If we think of the symbol Df as a map from $\mathbb{R}^n$ (or a subset of $\mathbb{R}^n$) to $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, then what is the derivative of the derivative? Think about that while we move to a more tractable problem. We have seen that, for a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, the partial derivative $D_j f$ can be thought of as a function from $\mathbb{R}^n$ to $\mathbb{R}$. Consequently, it would make sense to consider the partial derivatives of the functions $D_j f$. For example, suppose $f : \mathbb{R}^3 \longrightarrow \mathbb{R}$ is defined by $f(x, y, z) = x^2 y + e^{xz}$. Then, $D_1 f(x, y, z) = 2xy + ze^{xz}$. Then, $D_3(D_1 f)(x, y, z) = e^{xz} + zxe^{xz}$. Looking at this from another perspective, we next compute $D_3 f(x, y, z)$. We get $D_3 f(x, y, z) = xe^{xz}$ and $D_1(D_3 f)(x, y, z) = e^{xz} + xze^{xz}$. So, for this f, we see that

$$D_3(D_1 f)(x, y, z) = D_1(D_3 f)(x, y, z).$$

The functions $D_3(D_1 f)$ and $D_1(D_3 f)$ are examples of *mixed partial derivatives* of f.

For a general setting, consider a function $f = (f_1, f_2, \ldots, f_m)$ from an open set in $\mathbb{R}^n$ to $\mathbb{R}^m$. Then, we define the mixed partial derivative

$$D_{ij} f_k = D_i(D_j f_k)$$

assuming that the $D_j f_k$ has partial derivatives for all j, k. The question is, "Under what conditions will we have $D_{ij} f_k = D_{ji} f_k$ for all i, j?"

THEOREM 5.5.1. *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \longrightarrow \mathbb{R}$ be differentiable on U. Suppose that D_{ij} and D_{ji} exist and are continuous on U. Then $D_{ij} f(x) = D_{ji} f(x)$ for all $x \in U$.*

Proof. It is enough to prove the theorem for $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$. Let $x = (a, b) \in U$, and let h be small enough that $B_{2h}(x)$ is contained in U. We consider the second differences for computing the partial derivatives of f. Set

$$A(h) = \frac{1}{h^2} \left(f(a+h, b+h) - f(a, b+h) - f(a+h, b) + f(a, b) \right).$$

By the Mean Value Theorem, there exist ξ and ξ' between a and $a+h$ such that

$$f(a+h, b+h) - f(a, b+h) = h D_1 f(\xi, b+h)$$

and

$$f(a+h, b) - f(a, b) = h D_1 f(\xi', b).$$

The Mean Value Theorem also gives ξ'' between ξ and ξ' and η between b and $b+h$ such that

$$D_1 f(\xi, b+h) - D_1 f(\xi', b) = h D_{21} f(\xi'', \eta) = h A(h).$$

For the next step, we rewrite

$$A(h) = \frac{1}{h^2} \left(f(a+h, b+h) - f(a+h, b) - f(a, b+h) + f(a, b) \right)$$

and proceed similarly. From this we get $A(h) = D_{12} f(\xi^*, \eta^*)$ where ξ^* and η^* are obtained similarly. If we let h go to 0, the continuity of the mixed partials now implies the result.

EXERCISE 5.5.2. Let

$$f(x, y) = \begin{cases} 0 & \text{when } (x, y) = (0, 0), \\ \frac{x^3 y - x y^3}{x^2 + y^2} & \text{otherwise.} \end{cases}$$

Show that f is differentiable everywhere. Show that $D_{12} f(0, 0)$ and $D_{21} f(0, 0)$ exist, but $D_{12} f(0, 0) \neq D_{21} f(0, 0)$.

In one variable calculus, higher derivatives are defined by differentiating the derivative considered as a function on some set. In the present situation, where we have a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, the derivative is a map from $\mathbb{R}^n$ to the space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ of linear maps $\mathbb{R}^n \longrightarrow \mathbb{R}^m$. From our discussion of linear algebra in Appendix B we know that $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ can be identified

with the space $\mathbb{R}^{mn}$. We can then interpret the second derivative at a point as an element of

$$\mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)) \cong \mathcal{L}(\mathbb{R}^n, \mathbb{R}^{mn}).$$

This space has mn^2 dimensions, and the entries of the matrix representing the derivative are the partial derivatives $D_{ij}f_k$. We could continue this process and define the ℓ-th derivative, but this would not be particularly useful for this text and we confine ourselves to the first and second derivatives.

On the other hand, it will be useful to talk about all orders of differentiation for the partial derivatives. For example, if $f : \mathbb{R}^5 \longrightarrow \mathbb{R}$ has partial derivatives of order 3 on an open set $U \subseteq \mathbb{R}^n$, that is, $D_1^3 f(x)$, $D_2^2 D_4 f(x)$, $D_2 D_4 D_2 f(x)$, etc., all exist, we can consider the situation similar to that in Theorem 5.5.1 in which two of these are equal.

Definition 5.5.3. Let $U \subseteq \mathbb{R}^n$ be an open set and let $f : U \longrightarrow \mathbb{R}^m$. The function f is said to be in $C^k(U)$ (or C^k on U) if f has all continuous partial derivatives up to and including order k on U. The function f is said to be in $C^\infty(U)$ or *smooth* on U if f has all partial derivatives of all orders on U.

Exercise 5.5.4. Suppose $U \subseteq \mathbb{R}^n$ is open and $f : U \longrightarrow \mathbb{R}$ is C^k on U. Let $\ell \leq k$ and $i_1, i_2, \ldots, i_\ell$ be a collection of integers between 1 and n. Show that, for any $\sigma \in S_\ell$,

$$D_{i_1,i_2,\ldots,i_\ell} f = D_{\sigma(i_1),\sigma(i_2),\ldots,\sigma(i_\ell)} f.$$

If f is C^k on an open set U in $\mathbb{R}^n$, then every partial derivative of order k can be written in the form $D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n} f$, where the α_i are nonnegative integers and $\sum_{i=1}^n \alpha_i = k$. There is an interesting combinatorial problem that arises here.

Exercise 5.5.5.

(i) How many ways can one partition k into an ordered sum of n nonnegative summands?

(ii) How many ways can one partition k into n nonnegative summands if the order is ignored?

Exercise 5.5.6. This is an exercise in high school algebra. Consider the polynomial in n variables,

$$p(x_1, x_2, \ldots, x_n) = (x_1 + x_2 + \cdots + x_n)^k.$$

Show that, upon expanding this in monomials, the coefficient of $x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}$, where $\sum_{i=1}^n k_i = k$, is given by

$$\frac{k!}{k_1! k_2! \cdots k_n!}.$$

This number is called a *multinomial coefficient* and is denoted by $\binom{k}{k_1,k_2,\ldots,k_n}$.

EXERCISE 5.5.7. Recall Exercise 5.3.12. Let $z_0 \in \mathbb{C}$, and define $f : \mathbb{C} \longrightarrow \mathbb{C}$ by $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$, where $a_n \in \mathbb{C}$ for all n. Let r be the radius of convergence of this power series and suppose that $r > 0$. Show that f is infinitely differentiable on $B_r(z_0)$.

Recall that in the Mean Value Theorem for functions $f : U \longrightarrow \mathbb{R}$, where U is an open set in $\mathbb{R}^n$, we stated that for points $x, y \in U$, there is an element z on the line joining x to y such that $f(y) = f(x) + Df(z)(y - x)$. For this to hold, the line joining x and y must lie in U. If we assume that f is C^{k+1}, we would like to further expand f in terms of its partial derivatives of order j where $j \leq k + 1$. How would we expect such an expression to look? First, let us write $y = x + th$, where $h = (h_1, h_2, \ldots, h_n)$ is a unit vector in $\mathbb{R}^n$ and $t > 0$. Then, the directional derivative of f at x in the direction of the vector h is $D_h f(x) = \nabla f(x) \cdot h$. If we iterate this directional derivative r times, we get the expression

$$D_h^r f(x) = \sum_{\alpha_1 + \cdots + \alpha_n = r} \binom{r}{\alpha_1, \ldots, \alpha_n} h_1^{\alpha_1} \cdots h_n^{\alpha_n} D_1^{\alpha_1} \cdots D_n^{\alpha_n} f(x).$$

We can now write

$$f(y) = f(x) + \sum_{r=1}^{k} \frac{D_h^r f(x) t^r}{r!} + R_k(y),$$

where, as one might suspect,

$$R_k(y) = f(y) - f(x) - \sum_{r=1}^{k} \frac{D_h^r f(x) t^r}{r!}.$$

THEOREM 5.5.8 **(Taylor's theorem)**. *Let U be a convex open set in $\mathbb{R}^n$, and let $f : U \longrightarrow \mathbb{R}$ be a C^{k+1} function. If $x, y \in U$, then there exists $s \in \mathbb{R}$, $0 < s < t$, such that*

$$R_k(y) = \frac{D_h^{k+1} f(x + sh) s^{k+1}}{(k+1)!}.$$

This is, of course, the multivariable analog of Corollary 5.1.24.

Proof. Apply Corollary 5.1.24 to the function $F(a) = f(x + ah)$.

DEFINITION 5.5.9. The polynomial

$$f(x) + \sum_{r=1}^{k} \frac{D_h^r f(x) t^r}{r!}$$

is called the *Taylor polynomial of degree k* for the function f at x in the direction h.

The Taylor polynomial may be regarded as an approximation to f in a neighborhood of the point x in the direction h because the remainder term vanishes to order $k + 1$ at x and has a $(k + 1)!$ in the denominator.

EXERCISE 5.5.10. Compute the Taylor polynomial of degree 3 for the following functions at the specified point (in an arbitrary direction h):

(i) $f(x,y,z) = \frac{1}{xyz}$ at $(1,1,1)$.
(ii) $f(x,y,z) = e^{xy+xz+yz}$ at $(0,0,0)$.

EXERCISE 5.5.11. Let U be a convex open set in $\mathbb{R}^n$, let $x \in U$, let $h \in \mathbb{R}^n$ be a unit vector, and let $f : U \longrightarrow \mathbb{R}$ be a C^{k+1} function. Show that the Taylor polynomial of f at x is the best polynomial approximation to f at x by proving that if P is a polynomial of degree k such that

$$\lim_{t\to 0} \left| \frac{f(x+th) - P(t)}{t^k} \right| = 0,$$

then P is the Taylor polynomial of degree k of f at x in the direction h.

5.6. Hypersurfaces and Tangent Hyperplanes in $\mathbb{R}^n$

As stated in Section 5.1, the derivative of a function of one variable $f : \mathbb{R} \longrightarrow \mathbb{R}$ that is differentiable at a point $x = c$ is the slope of the tangent line to the curve $y = f(x)$ at $(c, f(c))$. The equation of the tangent line can be written $y = f(c) + f'(c)(x - c)$.

This situation may be regarded as a special case of a tangent line to a general curve in $\mathbb{R}^2$. The curve $y = f(x)$ may be considered as a level curve of the function $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ given by $F(x,y) = f(x) - y$, where $F(x,y) = 0$.

We now assume that $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is any C^1 function. Consider the curve $F(x,y) = k$ in $\mathbb{R}^2$. Let (x_0, y_0) be a point on this curve, and assume that the gradient $\nabla F(x_0, y_0) \neq (0,0)$. The tangent line to the curve at the point (x_0, y_0) is the line through the point perpendicular to $\nabla F(x_0, y_0)$. The equation of this line is then

$$D_1F(x_0,y_0)(x - x_0) + D_2F(x_0,y_0)(y - y_0) = 0.$$

EXERCISE 5.6.1. Show that in the case $F(x,y) = f(x) - y$, the equation of the tangent line is the same as that given above.

EXERCISE 5.6.2.

(i) Given the function $F(x,y) = x^2 + y^2$, consider the curve $F(x,y) = 3$. That is, $x^2 + y^2 = 3$. Find the equation of the tangent line at each point of this curve.
(ii) Given the function

$$F(x,y) = \begin{cases} x^2 \sin^2(1/y) & \text{if } y \neq 0, \\ 0 & \text{if } y = 0, \end{cases}$$

consider the curve $F(x,y) = 1$. At what points is F a C^1 function? At what points is the gradient vector nonzero? Find an equation for the tangent line at all points where the gradient vector is nonzero.

How does this generalize to N-dimensional space? We consider a C^1 function $F : \mathbb{R}^N \longrightarrow \mathbb{R}$.

DEFINITION 5.6.3. A *smooth hypersurface* in $\mathbb{R}^N$ is a set of points defined by the equation

$$F(x_1, x_2, \ldots, x_N) = k$$

with the property that $\nabla F \neq 0$ at each point in this set.

EXAMPLE 5.6.4.

(i) Let $F : \mathbb{R}^N \longrightarrow \mathbb{R}$ be defined by $F(x_1, x_2, \ldots, x_N) = x_1^2 + x_2^2 + \cdots + x_N^2$. Then the smooth hypersurface defined by $F(x_1, x_2, \ldots, x_N) = 1$ is the unit sphere in $\mathbb{R}^N$.

(ii) Let $F : \mathbb{R}^3 \longrightarrow \mathbb{R}$ be defined by $F(x, y, z) = x^2 + y^2 - z$. Then the equation $F(x, y, z) = 0$ gives a smooth hypersurface in $\mathbb{R}^3$ that is called a *paraboloid*.

DEFINITION 5.6.5. Let S be a smooth hypersurface in $\mathbb{R}^N$ defined by $F : \mathbb{R}^N \longrightarrow \mathbb{R}$, and let $c = (c_1, c_2, \ldots, c_N) \in S$. We define the *tangent hyperplane* to S at c to be the hyperplane (see Appendix B) through the point c normal to the vector $\nabla F(c)$, that is, the hyperplane $(x - c) \cdot \nabla F(c) = 0$.

EXERCISE 5.6.6. Let $F : \mathbb{R}^3 \longrightarrow \mathbb{R}$ be given by

$$F(x, y, z) = \begin{cases} \frac{1}{x} + \frac{1}{y} + \frac{1}{z} & \text{if } x \neq 0, y \neq 0, z \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) For what values of k does the equation $F(x, y, z) = k$ define a smooth hypersurface?

(ii) For those values of k, find the equation of the tangent hyperplane at every point of the hypersurface.

EXERCISE 5.6.7. Let $F : \mathbb{R}^4 \longrightarrow \mathbb{R}$ be given by $F(x, y, z, w) = x^2 \sin y + y^2 \sin z - zwe^{xy}$.

(i) For what values of k does the equation $F(x, y, z, w) = k$ define a smooth hypersurface?

(ii) For those values of k, find the equation of the tangent hyperplane at every point of the hypersurface.

In some cases the discussion above can be made simpler. Suppose that the hypersurface defined by $F(x_1, x_2, \ldots, x_N) = k$ may be realized as the graph of a C^1 function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, where $N = n + 1$. Specifically, suppose the surface can be defined by $x_{n+1} = f(x_1, x_2, \ldots, x_n)$. For example, we may fix such a function f and define

$$F(x_1, x_2, \ldots, x_n, x_{n+1}) = f(x_1, x_2, \ldots, x_n) - x_{n+1}.$$

A point C on this hypersurface can be written as $C = (c, f(c))$, where $c = (c_1, c_2, \ldots, c_n)$ is a point in $\mathbb{R}^n$. Using the ideas developed above, we note that the tangent hyperplane to this hypersurface at the point C is

defined by $(X - C) \cdot \nabla F(C) = 0$, where $X = (x_1, \ldots, x_n, x_{n+1})$. Writing this out in coordinates, we get

$$(x_1 - c_1)D_1 f(c) + (x_2 - c_2)D_2 f(c) + \cdots + (x_n - c_n)D_n f(c) - (x_{n+1} - f(c)) = 0,$$

which can be rearranged to give

$$x_{n+1} = f(c) + (x_1 - c_1)D_1 f(c) + (x_2 - c_2)D_2 f(c) + \cdots + (x_n - c_n)D_n f(c).$$

EXERCISE 5.6.8. Show that this tangent hyperplane is the *n-dimensional affine subspace* of $\mathbb{R}^{n+1}$ through the point $C = (c, f(c))$ spanned by the set of vectors

$$\{v_j = (0, 0, \ldots, 0, 0, 1, 0, 0, \ldots, 0, 0, D_j f(c)) \mid j = 1, 2, \ldots, n\}$$

where the 1 in the vector v_j occurs in the j-th place.

EXERCISE 5.6.9. Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a C^1 function, and let $S \subseteq \mathbb{R}^{n+1}$ be the graph of f. Fix a point $c \in \mathbb{R}^n$, and let T be the tangent hyperplane to S at the point $C = (c, f(c)) \in S$.

(i) For any $v \in \mathbb{R}^n$, define $\gamma_v : \mathbb{R} \longrightarrow \mathbb{R}^n$ by $\gamma_v(t) = tv + c$. Let $\phi_v : \mathbb{R} \longrightarrow \mathbb{R}^{n+1}$ be defined by $\phi_v(t) = f(\gamma_v(t))$. Show that $C + \phi_v'(0)$ lies in T.

(ii) Show that every vector $X \in T$ can be written in the form $X = C + V$, where $V = \phi_v'(0)$ for some $v \in \mathbb{R}^n$. Hence, every vector in the tangent hyperplane may be realized as the tangent vector to a curve in S.

5.7. Max-Min Problems

We now consider the problem of finding maximum and/or minimum values of a function on various subsets of $\mathbb{R}^n$ using properties of the derivative. The first observation is that, in order to discuss the notion of maximum or minimum, we need an order on the range of the function. Thus, we will restrict our attention to real-valued functions for the rest of this discussion. The theory develops naturally along two lines: the first being when the domain of f is an open subset in $\mathbb{R}^n$ and the second when the domain of f is a hypersurface in $\mathbb{R}^n$ or the intersection of several hypersurfaces. A typical example of how both of these aspects come into play is if we consider a closed, bounded set $B \subseteq \mathbb{R}^n$ and a continuous function $f : B \longrightarrow \mathbb{R}$. We know from our discussion of continuous functions on metric spaces that f will assume a maximum and minimum value on B. Suppose that the maximum value is assumed at a point p. We call $f(p)$ the *global maximum* of f on B. Similarly, if f assumes a minimum value at q, we call $f(q)$ the *global minimum* of f on B. Note that the global maximum and global minimum may be assumed at several different points in B, indeed, even at an infinite number of points as in the case of a constant function. We will also consider local maxima and minima of a function below.

If such a point p (or q) occurs in the interior of B and if f is differentiable at p, then we will see that $Df(p) = 0$. If p occurs in the interior of B but f is not differentiable at p, we must explore the behavior of the function in a

neighborhood of the point p using various estimation techniques. Finally, if p is on the boundary of B, then it may occur that the boundary of B can be realized as a hypersurface or the intersection of hypersurfaces, which leads to the use of the theory of Lagrange multipliers to determine the point p. Note the natural correspondence with the one-variable case, where maxima and minima can occur at three types of points: critical points ($f' = 0$), singular points (f' does not exist), and endpoints.

We begin the analysis by considering $f : U \longrightarrow \mathbb{R}$, where U is an open subset of $\mathbb{R}^n$.

DEFINITION 5.7.1. Let U be an open set in $\mathbb{R}^n$, and let $f : U \longrightarrow \mathbb{R}$. We say that f assumes a *local maximum* at a point $p \in U$ if there exists $r > 0$ such that $B_r(p) \subseteq U$ and $f(x) \leq f(p)$ for all $x \in B_r(p)$.

Similarly, we can define a *local minimum* of f at a point q in U.

THEOREM 5.7.2. *Let U be an open set in $\mathbb{R}^n$, and let $f : U \longrightarrow \mathbb{R}$. If f has a local maximum at p and if f is differentiable at p, then $Df(p) = 0$.*

Proof. Write $p = (p_1, p_2, \ldots, p_n)$. From the hypotheses of the theorem, it follows that, for each $j = 1, 2, \ldots, n$,

$$f_j(x) = f(p_1, p_2, \ldots, p_{j-1}, x, p_{j+1}, \ldots, p_n)$$

has a local maximum as a function of the single variable x at p_j. Hence $D_j f(p) = f_j'(p_j) = 0$. Since f is differentiable, this implies $Df(p) = 0$.

DEFINITION 5.7.3. As in the one-variable case, a point where Df vanishes is called a *critical point* of f.

We note that, as in the case of one variable, f need not have either a local maximum or a local minimum at a critical point.

EXAMPLE 5.7.4. The function $f(x, y) = xy$ has vanishing derivative at the origin but has neither a maximum nor minimum there, since f is positive in the first and third quadrants and negative in the second and fourth quadrants.

EXAMPLE 5.7.5. The following example illustrates the care that we must take in identifying critical points. Let

$$g(x, y) = \begin{cases} \frac{2xy^2}{x^2+y^4} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

This function has all partial derivatives equal to zero at the origin yet is not even continuous there.

EXERCISE 5.7.6. Let $p = (x_0, y_0)$ be a point in the plane, and let $ax + by + c = 0$ be the equation of a line in the plane. Verify that the distance from p to this line is given by

$$\frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$$

using the max-min technique discussed above. Do the same for a point and a hyperplane in $\mathbb{R}^n$.

EXERCISE 5.7.7. Let $p_j = (x_j, y_j)$, $j = 1, \ldots, m$, be m points in $\mathbb{R}^2$ with at least two distinct x_j's. Given a line $y = mx + b$, define $E(m, b) = \sum_{j=1}^{m} (y_j - (mx_j + b))^2$. Find the values of m and b that minimize this sum. For the values of m and b that minimize the function E, the line $y = mx + b$ is called the *ordinary least squares approximation* to the data $p_1, \ldots, p_m$.

EXERCISE 5.7.8. Let $p_j = (x_j, y_j)$, $j = 1, \ldots, m$, be m points in $\mathbb{R}^2$ with at least two distinct x_j's. Given a line ℓ with equation $ax + by + c = 0$, with a and b not both zero, denote by $d(p_j, \ell)$ the distance from the point p_j to the line ℓ. Consider the function $E(a, b, c) = \sum_{j=1}^{m} d(p_j, \ell)$. Find values of a, b, and c that minimize this function.

EXERCISE 5.7.9. Given a point $p = (x_0, y_0, z_0)$ with $x_0, y_0, z_0 > 0$, find an equation for a plane passing through this point that cuts off a tetrahedron of least volume in the first octant.

Once we have identified the critical points of a function f, we might then ask if there is a convenient way to determine whether f actually assumes a local maximum or minimum value at these points. One surefire way is to check the behavior of the function in a neighborhood of the critical point directly using inequalities. The next theorem shows that there is a test, similar to the second derivative test for functions of one variable, for determining whether a function assumes a maximum or minimum value at a critical point.

Let U be an open set in $\mathbb{R}^n$ and let $f : U \longrightarrow \mathbb{R}$ be twice differentiable at a point $x \in U$. We define the *Hessian* of f at x to be the *quadratic form* $H_x : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by

$$H_x(v) = [D(Df)(x)(v)](v).$$

EXERCISE 5.7.10. Show that, in terms of the $n \times n$ matrix of second-order partial derivatives,

$$A_x = \begin{pmatrix} D_{11}f(x) & D_{12}f(x) & \cdots & D_{1n}f(x) \\ D_{21}f(x) & D_{22}f(x) & \cdots & D_{2n}f(x) \\ \vdots & \vdots & \ddots & \vdots \\ D_{n1}f(x) & D_{n2}f(x) & \cdots & D_{nn}f(x) \end{pmatrix},$$

we can write $H_x(v) = {}^t v A_x v$.

REMARK 5.7.11. We say that the quadratic form H_x is *positive definite* if $H_x(v) \geq 0$ for all $v \in \mathbb{R}^n$ and $H_x(v) = 0$ iff $v = 0$. Negative definiteness is defined similarly.

THEOREM 5.7.12 **(Second derivative test for extrema)**. *Let U be an open set in $\mathbb{R}^n$, let $f : U \longrightarrow \mathbb{R}$ be a C^2 function, and let p be a critical point of f. If H_p is positive definite, then f assumes a local minimum at p, and if H_p is negative definite, then f assumes a local maximum at p.*

Proof. It is enough to prove the statement about minima, since we can replace f by $-f$ for maxima. We claim that if H_p is positive definite, then H_x is also positive definite for x in a neighborhood of p. To see this, let $m = \inf_{\|v\|=1} H_p(v)$. Then $H_p(v) \geq m\|v\|^2$ for any vector v, because v is the product of $\|v\|$ and a unit vector.

EXERCISE 5.7.13. Using the fact that the $D_{ij}f$, which form the coefficients of H, are continuous in x, show that for x sufficiently close to p we have $H_x(v) \geq \frac{1}{2}m\|v\|^2$ for any vector v.

This proves our claim.

Given any sufficiently small vector h, the Taylor polynomial for f at p, to second order, is

$$f(p+h) = f(p) + Df(p) \cdot h + \frac{1}{2}H_{p+th}(h)t^2$$

for some $t \in [0,1]$. Since p is a critical point, $Df(p) = 0$. By the estimate on H_x above, we have $f(p+h) \geq f(p)$, which proves that f assumes a local minimum at p.

EXERCISE 5.7.14. For each of the following functions $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$, find the critical points and compute the Hessian at each such point. Use the second derivative test to determine whether the critical points are local maxima or local minima, if possible. If the test does not apply, determine the nature of the critical point by other means.

(i) $f(x,y) = x^2 + y^2$.
(ii) $f(x,y) = x^2 - 2xy + y^2$.
(iii) $f(x,y) = x^2 - y^4$.

EXAMPLE 5.7.15. Consider the function $f(x,y) = xy$. The only critical point of f is $(0,0)$. The Hessian matrix at this point is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and the associated quadratic form is

$$H_{(0,0)}(u,v) = 2uv.$$

Thus, the second derivative test does not apply. Nonetheless, it is easy to see that f has no local extremum at $(0,0)$.

EXERCISE 5.7.16. Let $U \subseteq \mathbb{R}^2$ be an open set and let $f : U \longrightarrow \mathbb{R}$ be C^2. Let $p \in U$ be a critical point of f and suppose that the matrix of H_p has negative determinant. Show that f does not have a local extremum at p.

The critical points of the type in Exercise 5.7.16 are referred to as *saddle points*.

EXERCISE 5.7.17. Find the critical points of $f(x, y) = x^3 + 8y^3 - 6xy - 2$. For each, determine if it is a local maximum, local minimum, or saddle point, if possible.

EXERCISE 5.7.18. Let $p_0 = (x_0, y_0, z_0)$ be a point in the first octant in $\mathbb{R}^3$, that is, $x_0, y_0, z_0 > 0$. Suppose a plane through p_0 intersects the x-, y-, and z-axes at $p_1 = (a, 0, 0)$, $p_2 = (0, b, 0)$, and $p_3 = (0, 0, c)$, respectively, with $a, b, c > 0$. Find the plane such that the area of the triangle with vertices p_1, p_2, and p_3 is minimized.

5.8. Lagrange Multipliers

We now turn to the case where we wish to find the extreme values of a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ restricted to a hypersurface in $\mathbb{R}^n$ or the intersection of several hypersurfaces. We will first deal with the case of a single hypersurface S. The equation for the hypersurface S is sometimes referred to as a *constraint*, and the overall technique we develop here is called optimization with constraints.

In the case that S can be realized as the graph of a function $g : \mathbb{R}^{n-1} \longrightarrow \mathbb{R}$, we are reduced to the problem of finding extreme values of the function $h : \mathbb{R}^{n-1} \longrightarrow \mathbb{R}$ defined by

$$h(x_1, \ldots, x_{n-1}) = f(x_1, \ldots, x_{n-1}, g(x_1, \ldots, x_{n-1}))$$

on the open set $\mathbb{R}^{n-1}$, which we have previously done.

Even if the equation defining S cannot be solved explicitly for one of the variables, it may be still possible to describe S by an unconstrained set of $n - 1$ variables by parameterization. While we will not pursue a full discussion of the notion of parameterization here, we will give a typical example of this approach.

EXAMPLE 5.8.1. Let $S = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}$ be the unit sphere in $\mathbb{R}^3$. We wish to find the maxima and minima of the function $f(x, y, z) = x + y + z$ on S. We may parameterize S by using spherical coordinates as follows. Recall that, in general, spherical coordinates are given by

$$\begin{aligned} x &= \rho \cos\theta \sin\phi, \\ y &= \rho \sin\theta \sin\phi, \\ z &= \rho \cos\phi. \end{aligned}$$

In the specific case at hand, however, we have $\rho = 1$, so S is parameterized by the two unconstrained variables θ and ϕ.

Written in spherical coordinates, the function f becomes $f_0(\theta, \phi) = \cos\theta \sin\phi + \sin\theta \sin\phi + \cos\phi$. This parameterization causes problems at points where the Jacobian of the change of variables map above is singular (that is, has determinant zero), and this occurs exactly when $\rho = 0$ or

$\sin\phi = 0$. In particular, we cannot determine whether points with $\sin\phi = 0$ are critical with this parameterization.

We have

$$\begin{aligned}\frac{\partial f_0}{\partial \theta} &= -\sin\theta\sin\phi + \cos\theta\sin\phi = \sin\phi(\cos\theta - \sin\theta),\\ \frac{\partial f_0}{\partial \phi} &= \cos\theta\cos\phi + \sin\theta\cos\phi - \sin\phi = \cos\phi(\cos\theta + \sin\theta) - \sin\phi.\end{aligned}$$

Note that $\partial f_0/\partial\theta = 0$ when $\cos\theta = \sin\theta$ or $\sin\phi = 0$. The former occurs when $\cos\theta = \sin\theta = \pm\frac{\sqrt{2}}{2}$. In this case, $\cos\theta + \sin\theta = \pm\sqrt{2}$. So $\partial f_0/\partial\phi$ also vanishes when $\pm\sqrt{2}\cos\phi - \sin\phi = 0$, that is, $\tan\phi = \pm\sqrt{2}$.

We have thus shown that points with $\sin\theta = \cos\theta = \pm\frac{\sqrt{2}}{2}$ and $\tan\phi = \pm\sqrt{2}$ (where the signs are the same) are among the critical points of f.

Extreme value problems of the type discussed in Example 5.8.1 can be approached in a somewhat different fashion, by the method of *Lagrange multipliers*. The basic idea works as follows. Let S be the level surface of a C^1 function $g : \mathbb{R}^n \longrightarrow \mathbb{R}$. Moreover, assume that $\nabla g(x) \neq 0$ for all $x \in S$, so that S is a smooth hypersurface. For each $x \in S$, $\nabla g(x)$ is, by definition, normal to the tangent hyperplane to S at x. Now suppose that the function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, when restricted to S, has an extreme value at the point $a \in S$. Then, for any C^1 curve $\phi : \mathbb{R} \longrightarrow S$ with $\phi(0) = a$, the function $f \circ \phi : \mathbb{R} \longrightarrow \mathbb{R}$ has an extreme value at 0, and hence

$$0 = (f \circ \phi)'(0) = \nabla f(\phi(0)) \cdot \phi'(0) = \nabla f(a) \cdot \phi'(0).$$

In other words, $\nabla f(a)$ is normal to the vector tangent to the curve ϕ at the point a. Since this is true for any such curve ϕ, $\nabla f(a)$ is normal to the tangent hyperplane to S at a by Exercise 5.9.14 (see also Exercise 5.6.9 in case the constraint is given by the graph of a function). Since $\nabla g(a)$ is also normal to the tangent hyperplane to S at a, this implies that $\nabla f(a)$ is proportional to $\nabla g(a)$, so we can write $\nabla f(a) = \lambda \nabla g(a)$ for some $\lambda \in \mathbb{R}$. The scalar λ is known as a *Lagrange multiplier*. Before proving a theorem about Lagrange multipliers, let us study the example above using this new idea.

Example 5.8.2. Let $S = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}$ be the unit sphere in $\mathbb{R}^3$. We wish to find the maxima and minima of the function $f(x, y, z) = x + y + z$ on S. We observe first that $\nabla f(x, y, z) = (1, 1, 1)$. The surface S can be realized as the level set $g(x, y, z) = 0$ for the function $g(x, y, z) = x^2 + y^2 + z^2 - 1$, which has gradient $\nabla g(x, y, z) = (2x, 2y, 2z)$. To identify the potential points $a = (x_0, y_0, z_0)$ on the surface S where f attains an extreme value, we set up the Lagrange multiplier equation $\nabla f(a) = \lambda \nabla g(a)$, which becomes

$$(1, 1, 1) = \lambda(2x_0, 2y_0, 2z_0).$$

This gives us three equations in our four unknowns, namely,

$$2\lambda x_0 = 1,$$
$$2\lambda y_0 = 1,$$
$$2\lambda z_0 = 1.$$

But we must not forget that we have the original constraint, namely, that $g(a) = 0$, and hence, there is a fourth equation,

$$x_0^2 + y_0^2 + z_0^2 = 1.$$

Solving this system of four equations in four unknowns gives $\lambda = \pm\frac{\sqrt{3}}{2}$, which leads to $x_0 = \pm\frac{\sqrt{3}}{3}$, $y_0 = \pm\frac{\sqrt{3}}{3}$, $z_0 = \pm\frac{\sqrt{3}}{3}$. The point $a_1 = (+\frac{\sqrt{3}}{3}, +\frac{\sqrt{3}}{3}, +\frac{\sqrt{3}}{3})$ is the point where f attains its maximum value of $\sqrt{3}$ when restricted to S, and the point $a_2 = (-\frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{3})$ is the point where f attains its minimum value of $-\sqrt{3}$ when restricted to S.

Finally, it is worth noting that in this case, the precise value of λ is irrelevant and was merely used in an auxiliary fashion to help find the potential extrema.

Let us confirm this method with a theorem.

THEOREM 5.8.3. *Let U be an open set in $\mathbb{R}^n$ and let $f : U \longrightarrow \mathbb{R}$ be C^1. Let $g : U \longrightarrow \mathbb{R}$ be C^1 and suppose that $\nabla g(x) \neq 0$ for all $x \in U$. Let S be the smooth hypersurface defined by $g(x) = 0$. If $f|_S$ has a local maximum or minimum at a point $a \in S$, then there exists $\lambda \in \mathbb{R}$ such that $\nabla f(a) = \lambda \nabla g(a)$.*

Proof. By Exercise 5.9.14 in the next section, every vector in the tangent hyperplane to S at a may be realized as the tangent vector $\phi'(0)$ to some C^1 curve $\phi : \mathbb{R} \longrightarrow S$ with $\phi(0) = a$. By the discussion above, $\nabla f(a)$ and $\nabla g(a)$ both lie in the affine space normal to the tangent hyperplane. By Exercise 5.9.13, which you will also do in the next section, this space is one-dimensional, so we are done.

EXAMPLE 5.8.4. Let us use the method of Lagrange multipliers to determine the points on the ellipse $x^2+4y^2 = 4$ that are closest to and farthest from the point $(1, 0)$. The square of the distance from the point (x, y) to this point is given by $f(x, y) = (x - 1)^2 + y^2$ and we wish to optimize this function subject to the constraint

$$g(x, y) = x^2 + 4y^2 - 4 = 0.$$

We have $\nabla f(x, y) = (2(x - 1), 2y)$ and $\nabla g(x, y) = (2x, 8y)$, so we consider the equations

$$\begin{aligned} 2(x - 1) &= \lambda(2x), \\ 2y &= \lambda(8y). \end{aligned}$$

Is is easy to see that the only points on the ellipse satisfying these equations for some λ are $(\pm 2, 0)$. Plugging into f, we conclude that $(2, 0)$ is the nearest point and $(-2, 0)$ is the farthest point.

EXERCISE 5.8.5. Let $S = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\}$ be the unit sphere in $\mathbb{R}^3$. Find the maxima and minima of the function $f(x, y, z) = x^3 + y^3 + z^3$ on S.

EXERCISE 5.8.6. Consider the function $P : (0, \infty) \times (0, \infty) \longrightarrow \mathbb{R}$ given by $P(L, C) = \alpha L^a C^b$, where α, a, b are positive constants and $a + b = 1$. Let $R : (0, \infty) \times (0, \infty) \longrightarrow \mathbb{R}$ be given by $R(L, C) = \beta_1 L + \beta_2 C$ for positive constants β_1 and β_2.

(i) Maximize P subject to the constraint $R(L, C) = \kappa_1$, where κ_1 is a positive constant.
(ii) Minimize R subject to the constraint $P(L, C) = \kappa_2$, where κ_2 is a positive constant.

In economics, the function P is known as the Cobb-Douglas production function.

EXERCISE 5.8.7. Let $x_1, x_2, \ldots, x_n$ be positive real numbers. Prove the *arithmetic-geometric mean inequality*,

$$(x_1 x_2 \cdots x_n)^{\frac{1}{n}} \leq \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Hint: Consider the function $f(x_1, x_2, \ldots, x_n) = \frac{x_1 + x_2 + \cdots + x_n}{n}$ subject to the constraint $x_1 x_2 \cdots x_n = c$, where c is a constant.

EXERCISE 5.8.8. If a triangle has side lengths x, y, and z, so its perimeter is $2s = x + y + z$, its area A satisfies $A^2 = s(s - x)(s - y)(s - z)$. Show that, among all triangles with given perimeter, an equilateral triangle has the largest area.

There is also a max-min theorem with several constraints.

THEOREM 5.8.9. *Let U be an open set in $\mathbb{R}^n$, let $f : U \longrightarrow \mathbb{R}$ be C^1, and let $g_1, g_2, \ldots, g_m : U \longrightarrow \mathbb{R}$ be C^1, where $m < n$. Assume that $\{\nabla g_1(x), \ldots, \nabla g_m(x)\}$ is a linearly independent set for all $x \in U$, and let S be the intersection of the hypersurfaces S_i defined by $g_i(x) = 0$. If $f|_S$ has a local maximum or minimum at a point $a \in S$, then $\nabla f(a)$ is a linear combination of $\nabla g_1(a), \nabla g_2(a), \ldots, \nabla g_m(a)$.*

If $\nabla f(a) = \lambda_1 \nabla g_1(a) + \lambda_2 \nabla g_2(a) + \cdots + \lambda_m \nabla g_m(a)$ as in the theorem, the scalars $\lambda_1, \lambda_2, \ldots, \lambda_m$ are called *Lagrange multipliers*. We do not prove this theorem here, but we present an example and some exercises to illustrate the theory.

EXAMPLE 5.8.10. Given the line defined by $P(x, y) = y - (mx + k) = 0$ and an ellipse defined by $E(x, y) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0$, we wish to find the

minimum distance between a point on the line and a point on the ellipse. Equivalently, we minimize the square distance

$$d(x_1, y_1, x_2, y_2) = (x_1 - x_2)^2 + (y_1 - y_2)^2$$

subject to the constraints

$$g_1(x_1, y_1, x_2, y_2) = 0 \quad \text{and} \quad g_2(x_1, y_1, x_2, y_2) = 0$$

where

$$g_1(x_1, y_1, x_2, y_2) = P(x_1, y_1)$$

and

$$g_2(x_1, y_1, x_2, y_2) = E(x_2, y_2).$$

We assume that the line does not intersect the ellipse. The reader should verify that this fact holds when $|k| > |b|$ and $m^2 < \frac{k^2 - b^2}{a^2}$. These conditions also emerge from the solution below.

We have

$$\begin{aligned} \nabla d &= (2(x_1 - x_2), 2(y_1 - y_2), -2(x_1 - x_2), -2(y_1 - y_2)), \\ \nabla g_1 &= (-m, 1, 0, 0), \\ \nabla g_2 &= \left(0, 0, \frac{2x_2}{a^2}, \frac{2y_2}{b^2}\right). \end{aligned}$$

We first note that ∇g_1 and ∇g_2 are everywhere linearly independent, so by Theorem 5.8.9, if (x_1, x_2, y_1, y_2) is a maximum or minimum value for d subject to the constraints g_1 and g_2, then the following system of six equations in six unknowns must be satisfied:

$$\begin{aligned} 2(x_1 - x_2) &= -\lambda_1 m, \\ 2(y_1 - y_2) &= \lambda_1, \\ -2(x_1 - x_2) &= \lambda_2 \frac{2x_2}{a^2}, \\ -2(y_1 - y_2) &= \lambda_2 \frac{2y_2}{b^2}, \\ y_1 &= mx_1 + k, \\ \frac{x_2^2}{a^2} + \frac{y_2^2}{b^2} - 1 &= 0. \end{aligned}$$

EXERCISE 5.8.11. Solve the above system of equations and find the minimum distance from the line to the ellipse.

EXERCISE 5.8.12. Consider the plane defined by $P(x, y, z) = Ax + By + Cz + D = 0$ and the ellipsoid defined by $E(x, y, z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 = 0$.

(i) Find conditions on A, B, C, D, a, b, c such that the plane and the ellipsoid do not intersect.

(ii) Find the minimum distance between the plane and the ellipsoid when they do not intersect.

EXERCISE 5.8.13. Let v and w be vectors in $\mathbb{R}^n$. Find the maximum and minimum values of $f(v, w) = v \cdot w$ subject to the constraints $\|v\| = \|w\| = 1$.

EXERCISE 5.8.14. Consider two nonparallel planes in $\mathbb{R}^3$. Find the point on their line of intersection closest to the origin in $\mathbb{R}^3$.

EXERCISE 5.8.15. In the situation of Theorem 5.8.9, what happens if the number of constraints exceeds the number of variables, that is, if $m \geq n$?

5.9. The Implicit and Inverse Function Theorems

Let n and m be positive integers, and let f be a function from $\mathbb{R}^m \times \mathbb{R}^n$ to $\mathbb{R}^m$. We attack the problem of determining a set of conditions under which we can solve $f = 0$ explicitly for the first m variables, which we denote by y's, in terms of the remaining n variables, which we denote by x's. Thus, if we write f in the form $f(y_1, y_2, \ldots, y_m; x_1, x_2, \ldots, x_n)$ as a function of $m+n$ variables, we would like to produce functions $\phi_1, \phi_2, \ldots, \phi_m$ from an open set in $\mathbb{R}^n$ to $\mathbb{R}$ such that, on some open set in $\mathbb{R}^m \times \mathbb{R}^n$, the assignment $y_j = \phi_j(x_1, x_2, \ldots, x_n)$ solves the equation $f(y_1, y_2, \ldots, y_m; x_1, x_2, \ldots, x_n) = 0$.

Of course, we can expect some conditions on differentiability, nonvanishing of the determinant of a Jacobian matrix, and other properties. We begin by considering the case when $m = 1$ and n is arbitrary. Here, the proof involves only simple results from basic calculus. This development can be completed by induction on m, but the techniques are somewhat tedious.

EXAMPLE 5.9.1. Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be defined by $f(y, x) = y^2 + x^2 - 1$. Here, the equation $f(y, x) = 0$ gives us the unit circle in $\mathbb{R}^2$. As is ordinarily discussed in elementary calculus courses, our analysis produces two functions that serve our purpose here. We wish to solve for y in terms of x, and this can be done with either of the equations $y = \sqrt{1 - x^2}$, for $-1 \leq x \leq 1$, or $y = -\sqrt{1 - x^2}$, for $-1 \leq x \leq 1$. Note that, in either case, we have $f(y, x) = 0$. Also, $\partial f / \partial y = 2y \neq 0$ when $y \neq 0$. This condition about the nonvanishing of the derivative is one that will be required in the next theorem.

THEOREM 5.9.2. *Let f be a C^1 function from an open set in $\mathbb{R} \times \mathbb{R}^n$ to $\mathbb{R}$. Let (y_0, x_0) be a point in this open set such that $f(y_0, x_0) = 0$ and*

$$\frac{\partial f}{\partial y}(y_0, x_0) \neq 0.$$

Then there exist open sets $V \subseteq \mathbb{R}$ and $U \subseteq \mathbb{R}^n$, such that $(y_0, x_0) \in V \times U$, and for every $x \in U$, there exists a unique $y = \phi(x)$ in V such that $f(\phi(x), x) = 0$ and $\phi : U \longrightarrow \mathbb{R}$ is C^1 on U. Furthermore,

$$\frac{\partial \phi}{\partial x_j}(x) = -\frac{\frac{\partial f}{\partial x_j}(\phi(x), x)}{\frac{\partial f}{\partial y}(\phi(x), x)}.$$

Proof. We can assume that $\frac{\partial f}{\partial y}(y_0, x_0) > 0$. Since f is C^1, there is an open set $U' \subseteq \mathbb{R}^n$ containing x_0 and an open interval $V = (y_0 - \varepsilon, y_0 + \varepsilon) \subseteq \mathbb{R}$ such that $\frac{\partial f}{\partial y}(y, x) > 0$ for $x \in U', y \in V$. Then $f(y_0 + \varepsilon, x_0) > 0$, and $f(y_0 - \varepsilon, x_0) < 0$. By continuity of f, there exists an open set $U \subseteq U' \subseteq \mathbb{R}^n$ containing x_0 such that, for all $x \in U$, $f(y_0 - \varepsilon, x) < 0$ and $f(y_0 + \varepsilon, x) > 0$.

Fix $x \in U$. By the Intermediate Value Theorem, there exists $y \in V$ such that $f(y, x) = 0$. The function $g(y) = f(y, x)$ satisfies $g'(y) = \frac{\partial f}{\partial y}(y, x) > 0$ for $y \in V$, so by Rolle's theorem, the value of y for which $f(y, x) = 0$ is unique. We set $\phi(x) = y$.

The continuity of ϕ at x_0 follows from the fact that we can choose ε in the above construction to be arbitrarily small. This same argument holds for any $x \in U$, which proves the continuity of ϕ.

Since $f(\phi(x), x) = 0$, formally, from the chain rule, we get

$$0 = \frac{\partial}{\partial x_j}[f(\phi(x), x)] = \frac{\partial f}{\partial y}(\phi(x), x)\frac{\partial \phi}{\partial x_j}(x) + \frac{\partial f}{\partial x_j}(\phi(x), x).$$

Thus,

$$\frac{\partial \phi}{\partial x_j}(x) = -\frac{\frac{\partial f}{\partial x_j}(\phi(x), x)}{\frac{\partial f}{\partial y}(\phi(x), x)}.$$

Using the formal expression for the derivative given above, we can write the difference quotient for the derivative of ϕ, subtract this expression, and show that the limit of the difference equals zero.

A continuation of this proof to yield the Implicit Function Theorem for general m is outlined in Osgood [**19**]. If the reader enjoys a stiff climb over rocks and across streams, he or she might wish to pursue this proof. As an alternative, you might wish to consult Folland's Appendix in [**5**].

We now make an about-face and move directly to a proof of the Inverse Function Theorem, a common approach in current mathematics texts. We first review the single-variable case, to remind the reader of the nature of the result.

Suppose that U is an open set in $\mathbb{R}$ and that $f : U \longrightarrow \mathbb{R}$ is C^1. Take a point $x_0 \in U$. We saw earlier (see Corollary 5.1.15) that if $f'(x_0) \neq 0$, then f is monotonic in an open interval I around x_0. This, of course, implies that f is one-to-one on I. Moreover, $f(I)$ is an open interval J contained in $\mathbb{R}$, and $f^{-1} : J \longrightarrow I$ is C^1 and $(f^{-1})'(y) = (f'(f^{-1}(y)))^{-1}$. It is worth remarking at this point that this one-variable theorem requires the continuity of the derivative. See, for example, Exercise 5.1.20.

The *Inverse Function Theorem* is the generalization of this result to functions $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$. Essentially, the theorem says that if such an f is C^1 and has a nonsingular derivative at a point x_0, then f is invertible in some neighborhood of x_0 and f^{-1} is also C^1. We approach this through a sequence of lemmas and corollaries.

LEMMA 5.9.3. *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \longrightarrow \mathbb{R}^n$ be C^1. Take $x_0 \in U$ and suppose that $Df(x_0)$ is nonsingular. Then there exists a neighborhood W of x_0 and a constant $c > 0$ such that*

$$\|f(y) - f(x)\| \geq c\|y - x\| \quad \textit{for all } x, y \in W.$$

Proof. For any nonsingular linear transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, we know that $\|T(y) - T(x)\| \leq \|T\|\|y - x\|$. It follows immediately that

$$\|T(y) - T(x)\| \geq \|T^{-1}\|^{-1}\|y - x\|.$$

Take $c = \|Df(x_0)^{-1}\|^{-1}/2$. Suppose that $f_1, f_2, \ldots, f_n$ are the component functions of f. Of course, these are C^1, so there exists a convex neighborhood W of x_0 such that $\|Df_i(y) - Df_i(x_0)\| \leq c/n$ for $y \in W$ and all i. Now the Mean Value Theorem implies that, given $x, y \in W$, there exists a point ξ_i on the line segment joining x and y such that

$$f_i(y) - f_i(x) = Df_i(\xi_i)(y - x).$$

Consequently, for $x, y \in W$ and each i, we have

$$\|f_i(y) - f_i(x) - Df_i(x_0)(y - x)\| \leq \frac{c}{n}\|y - x\|.$$

It follows immediately that

$$\|f(y) - f(x) - Df(x_0)(y - x)\| \leq c\|y - x\|.$$

Now, using the triangle inequality, we get

$$\|f(y) - f(x)\| \geq c\|y - x\| \quad \text{for } x, y \in W.$$

COROLLARY 5.9.4. *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \longrightarrow \mathbb{R}^n$ be C^1. Take $x_0 \in U$ and suppose that $Df(x_0)$ is nonsingular. Then there exists a neighborhood W of x_0 such that $f|_W$ is one-to-one.*

Proof. Exercise.

COROLLARY 5.9.5. *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \longrightarrow \mathbb{R}^n$ be C^1. Take $x_0 \in U$ and suppose that $Df(x_0)$ is nonsingular. Then there exists a neighborhood V of x_0 such that $f(V)$ is open and $f|_V : V \longrightarrow f(V)$ is a homeomorphism.*

Proof. Using the previous lemma and the fact that f is C^1, we can pick a neighborhood W of x_0 such that $\overline{W} \subseteq U$ and for some constant c, $\|f(y) - f(x)\| \geq c\|y - x\|$ for all $x, y \in \overline{W}$, and finally $Df(x)$ is nonsingular for $x \in \overline{W}$. Let V be any open ball contained in W and let $S = \partial\overline{V}$. Given a point $x \in V$ with $y = f(x) \notin f(S)$, since $f(S)$ is compact, the distance from y to $f(S)$, which we denote by d, is greater than zero. To show that $f(V)$ is open, we establish that $f(V)$ contains $B_{d/2}(y)$. To see this, take $z \in B_{d/2}(y)$. Then $\|z - y\| < d/2$. Moreover, the distance from z to $f(\overline{V})$, which equals

$$\inf_{x \in \overline{V}} \|z - f(x)\|,$$

is less than $d/2$. Since the distance from y to $f(S)$ is equal to d, it follows from the triangle inequality that the distance from z to $f(S)$ is greater than $d/2$. For $x \in \overline{V}$, we define the function

$$g(x) = \|z - f(x)\|^2 = \sum_{i=1}^{n} (z_i - f_i(x))^2.$$

We want to minimize this function. By compactness, there exists $x_1 \in \overline{V}$ such that

$$\|z - f(x_1)\|^2 = d(z, f(\overline{V}))^2.$$

From the previous inequalities, it follows that $x_1 \notin S$, so that $x_1 \in V$. So the minimum of g occurs at x_1. This implies that

$$0 = D_j g(x_1) = -2 \sum_{i=1}^{n} (z_i - f_i(x_1)) D_j f_i(x_1).$$

It follows immediately that $Df(x_1)(z - f(x_1)) = 0$, and since $Df(x_1)$ is nonsingular, $z = f(x_1)$. Since f is invertible on V by Corollary 5.9.4, f is a homeomorphism.

THEOREM 5.9.6 **(Inverse Function Theorem)**. *Let U be an open set in $\mathbb{R}^n$ and let $f : U \longrightarrow \mathbb{R}^n$ be C^1. Let $x_0 \in U$ be such that $Df(x_0)$ is nonsingular. Then there exists a neighborhood V of x_0 such that*

(i) *$f : V \longrightarrow f(V)$ is a bijection;*
(ii) *$f(V)$ is an open set in $\mathbb{R}^n$;*
(iii) *$f^{-1} : f(V) \longrightarrow V$ is C^1 and $Df^{-1}(f(x)) = (Df(x))^{-1}$ for $x \in V$.*

Proof. The first two statements follow from Corollary 5.9.4 and Corollary 5.9.5.

Now where are we? We have proved that $f : V \longrightarrow f(V)$ is one-to-one and that $f(V)$ is open. We consider the map $f^{-1} : f(V) \longrightarrow V$. By the last corollary, this map is continuous. We want to prove that it is C^1. A formal computation using the chain rule shows that $Df^{-1}(f(x))$, if it exists, is equal to $(Df(x))^{-1}$. To complete the proof, we take $y = f(x) \in f(V)$ and consider the difference quotient

$$\frac{\|f^{-1}(z) - f^{-1}(y) - Df(x)^{-1}(z - y)\|}{\|z - y\|}.$$

We can write this as

$$\frac{\|Df(x)^{-1}(Df(x)(f^{-1}(z) - f^{-1}(y)) - (z - y))\|}{\|z - y\|},$$

which is less than or equal to

$$\|Df(x)^{-1}\| \frac{\|Df(x)(f^{-1}(z) - f^{-1}(y)) - (z - y)\|}{\|z - y\|}.$$

Also, using Lemma 5.9.3, this is less than or equal to

$$\frac{\|Df(x)^{-1}\|}{c}\frac{\|Df(x)(f^{-1}(z)-x)-(z-f(x))\|}{\|f^{-1}(z)-x\|}.$$

Finally, since f^{-1} is continuous, this last expression goes to zero as z goes to y.

The function Df^{-1} is continuous because f is C^1 and, by Exercise 5.9.7 below, the map $A \mapsto A^{-1}$ is continuous from $GL_n(\mathbb{R}) \longrightarrow GL_n(\mathbb{R})$.

EXERCISE 5.9.7. Prove that the map $f : GL_n(\mathbb{R}) \longrightarrow GL_n(\mathbb{R})$ given by $f(A) = A^{-1}$ is continuous.

We now use the Inverse Function Theorem to outline a proof of the Implicit Function Theorem in a sequence of exercises.

THEOREM 5.9.8 **(Implicit Function Theorem)**. *Let f be a C^1 function from an open set in $\mathbb{R}^m \times \mathbb{R}^n$ to $\mathbb{R}^m$. Let (y_0, x_0) be a point in this open set such that $f(y_0, x_0) = 0$ and the matrix*

$$L = \left(\frac{\partial f_i}{\partial y_j}(y_0, x_0)\right)_{i,j=1,\ldots,m}$$

is nonsingular. Then there exist open sets $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ such that $(y_0, x_0) \in V \times U$, and to every $x \in U$, there exists a unique $y = \phi(x)$ in V such that $f(\phi(x), x) = 0$ and ϕ is C^1 on U.

Proof. To begin, we replace f by the function $F : \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}^m \times \mathbb{R}^n$ defined by $F(y, x) = (f(y, x), x)$.

EXERCISE 5.9.9. Show that $\det DF(y_0, x_0) = \det L$.

By applying the Inverse Function Theorem to F, we obtain neighborhoods $U' \subseteq \mathbb{R}^n$ of x_0 and $V \subseteq \mathbb{R}^m$ of y_0 such that F has a C^1 inverse on the open set $W = F(V \times U')$. Set $U = \{x \in U' \mid (0, x) \in W\}$.

EXERCISE 5.9.10. Show that there exists a C^1 function $\Phi : W \longrightarrow \mathbb{R}^m$ such that $F^{-1}(y, x) = (\Phi(y, x), x)$ on W.

EXERCISE 5.9.11. Show that $f(\Phi(y, x)) = y$.

EXERCISE 5.9.12. Let $\phi(x) = \Phi(0, x)$. Show that ϕ satisfies the conclusion of the theorem.

EXERCISE 5.9.13. Let $F : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ be a C^1 function such that $\nabla F(x) \neq 0$ for all x with $F(x) = 0$. Let S be the smooth hypersurface defined by $F(x) = 0$. Show that the tangent hyperplane to S at every $x_0 \in S$ is spanned by n linearly independent vectors. *Hint*: Use the Implicit Function Theorem, and apply Exercise 5.6.8.

EXERCISE 5.9.14. Let $F : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ be a C^1 function such that $\nabla F(x) \neq 0$ for all x with $F(x) = 0$. Let S be the smooth hypersurface defined by $F(x) = 0$. Show that every vector in the tangent hyperplane to S at $x_0 \in S$ is the tangent vector to some C^1 curve in S through x_0. *Hint*: Use the Implicit Function Theorem, and apply Exercise 5.6.9.

EXERCISE 5.9.15. Let $f : \mathbb{R}^3 \setminus \{(0,0,0)\} \longrightarrow \mathbb{R}^3 \setminus \{(0,0,0)\}$ be given by

$$f(x,y,z) = \left(\frac{x}{x^2+y^2+z^2}, \frac{y}{x^2+y^2+z^2}, \frac{z}{x^2+y^2+z^2} \right).$$

Show that f is locally invertible at every point in $\mathbb{R}^3 \setminus \{(0,0,0)\}$. Find an explicit formula for f^{-1}.

EXERCISE 5.9.16. Consider the equations

$$ab^2 + cde + a^2 d = 3 \quad \text{and} \quad ace^3 + 2bd - b^2 e^2 = 2.$$

Determine which pairs of variables can be solved for in terms of the other three near the point $(a,b,c,d,e) = (1,1,1,1,1)$.

We now present a result, Lemma 5.9.19 related to the previous theorems that will be of assistance in the change of variables theorem for multiple integrals. Given a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, we want to define the best linear approximation to f at a point using the Taylor polynomial of a real-valued function on $\mathbb{R}^n$. Let $U \subseteq \mathbb{R}^n$ be an open set and let $f : U \longrightarrow \mathbb{R}^n$ be a C^1 function. For $y \in U$, define a function $T^y(x)$ by $T^y(x) = (T_1^y(x), T_2^y(x), \ldots, T_n^y(x))$ where

$$T_j^y(x) = f_j(y) + \sum_{k=1}^{n} D_k f_j(y)(x_k - y_k).$$

This is the first-order Taylor polynomial of f_j in the direction of the unit-vector $(x-y)/\|x-y\|$ evaluated at $t = \|x-y\|$ in the terminology of Definition 5.5.9.

LEMMA 5.9.17. *Let $K \subseteq U$ be compact. Then*

$$\lim_{x \to y} \frac{\|T^y(x) - f(x)\|}{\|x-y\|} = 0$$

uniformly for $y \in K$.

Proof. It is enough to prove the analogous statement for each component function separately. By the Mean Value Theorem, there exists $\xi \in [0,1]$ such that

$$\begin{aligned} |T_j^y(x) - f_j(x)| &= \left| \sum_{k=1}^{n} (D_k f_j(y) - D_k f_j(y + \xi(x-y)))(x_k - y_k) \right| \\ &\leq \|x-y\| \|Df_j(y) - Df_j(y + \xi(x-y))\| \end{aligned}$$

by the Cauchy-Schwarz inequality. Dividing both sides by $\|x-y\|$, we get

$$\frac{|T_j^y(x)-f_j(x)|}{\|x-y\|} \leq \|Df_j(y)-Df_j(y+\xi(x-y))\|.$$

The right-hand side goes to zero as $x \to y$ uniformly for $y \in K$.

DEFINITION 5.9.18. We define a *rectangle* in $\mathbb{R}^n$ to be a set of the form $R = I_1 \times I_2 \times \cdots \times I_n$, where the I_i, $i = 1, \ldots, n$, are bounded intervals in $\mathbb{R}$. If the intervals are all open, we call R an *open rectangle*, and if the intervals are all closed, we call R a *closed rectangle*. In the particular case when the intervals are of the form $I_i = [a_i, b_i)$, we refer to R as a *half-open rectangle*.

LEMMA 5.9.19. *Suppose that $f : U \longrightarrow f(U)$ is C^1, one-to-one onto its image and has C^1 inverse. Let $R \subseteq U \subseteq \mathbb{R}^n$ be a rectangle with center y. For $\lambda > 0$, we denote by λR the rectangle with center y similar to R with sides scaled by λ. For $\varepsilon > 0$ and $h > 0$, we define*

$$\begin{aligned} R_1 &= (1-\varepsilon)hR, \\ R_2 &= hR, \\ R_3 &= (1+\varepsilon)hR. \end{aligned}$$

Then, for each compact set $K \subseteq U$ containing y, there exists a number $h_0(K) > 0$, independent of y, such that if $0 < h < h_0(K)$ and $R_2 \subseteq U$, then

$$T^y(R_1) \subseteq f(R_2) \subseteq T^y(R_3).$$

Proof. Since T^y has a continuously differentiable inverse, there exists a constant C such that

$$\|x-z\| \leq C\|T^y(x)-T^y(z)\|$$

for all $x, y, z \in K$. If we apply this estimate to the previous lemma, we have

$$\frac{\|T^y(x)-f(x)\|}{\|T^y(x)-f(y)\|} \longrightarrow 0$$

as $x \to y$ uniformly in y. If $x \in \partial R_2$, then

$$\frac{\|T^y(x)-f(x)\|}{\|T^y(x)-f(y)\|} \longrightarrow 0$$

as $h \to 0$, so there exists a constant $h_0(K) > 0$ such that, for $0 < h < h_0(K)$,

$$f(\partial R_2) \subseteq T^y(R_3) \setminus \overline{T^y(R_1)}.$$

It follows that $f(R_2) \subseteq T^y(R_3)$.

Since $f(R_2)$ has nonempty interior and contains $f(y) = T^y(y)$, there exists some $h' > 0$ such that $T^y(h'R) \subseteq f(R_2)$. Let

$$h'' = \sup\{h' \mid T^y(h'R) \subseteq f(R_2)\}.$$

Then $T^y(h''R) \cap f(\partial R_2) \neq \varnothing$, and hence $h'' > (1-\varepsilon)h$. Thus, $T^y(R_1) \subseteq f(R_2)$.

5.10. Independent Projects

5.10.1. Lebesgue's Theorem on the Differentiability of a Monotone Function.

DEFINITION 5.10.1. A set $A \subseteq \mathbb{R}$ has measure zero if, for any $\varepsilon > 0$, there exists a countable collection of intervals, $\{I_i = (a_i, b_i)\}_{i\in\mathbb{N}}$, such that $A \subseteq \bigcup_{i\in\mathbb{N}} I_i$ and $\sum_{i\in\mathbb{N}}(b_i - a_i) < \varepsilon$.

EXERCISE 5.10.2. Let $A \subseteq \mathbb{R}$. Show that A has measure zero if and only if there exists a countable collection of open intervals $\{I_i = (a_i, b_i)\}_{i\in\mathbb{N}}$ such that

(i) $\sum_{i\in\mathbb{N}}(b_i - a_i)$ is finite and
(ii) each $x \in A$ is contained in infinitely many I_i.

DEFINITION 5.10.3. If $P(x)$ is a property of the real number x, we say that P holds *almost everywhere* if $\{x \in \mathbb{R} \mid P(x) \text{ is false}\}$ has measure zero.

This definition allows us to state the main result of this project.

THEOREM 5.10.4 **(Lebesgue's theorem)**. *If $f : [a, b] \longrightarrow \mathbb{R}$ is a monotonic function, then f is differentiable almost everywhere.*

It is sufficient to prove the theorem for monotonic increasing functions. The following definitions and exercises will assist us in proving Lebesgue's theorem.

DEFINITION 5.10.5. Let $f : [a, b] \longrightarrow \mathbb{R}$. If, for any $c \in [a, b)$, the limit

$$\lim_{x\to c^+} f(x)$$

exists and, for any $c \in (a, b]$, the limit

$$\lim_{x\to c^-} f(x)$$

exists, then we say that f is *nearly continuous*.

EXERCISE 5.10.6. If $f : [a, b] \longrightarrow \mathbb{R}$ is monotone increasing, then f is nearly continuous.

EXERCISE 5.10.7. If $f : [a, b] \longrightarrow \mathbb{R}$ is monotone increasing, then f is continuous almost everywhere.

We will also make use of the following notation.

DEFINITION 5.10.8. If $f : [a, b] \longrightarrow \mathbb{R}$ is nearly continuous, then we define the following:

- $f^+(x) : [a, b] \longrightarrow \mathbb{R}$ is defined by $f^+(b) = f(b)$ and $f^+(c) = \lim_{x\to c^+} f(x)$ for $c \in [a, b)$.
- $f^-(x) : [a, b] \longrightarrow \mathbb{R}$ is defined by $f^-(a) = f(a)$ and $f^-(c) = \lim_{x\to c^-} f(x)$ for $c \in (a, b]$.

For a monotone increasing function $f : [a, b] \longrightarrow \mathbb{R}$, we also define the following functions from (a, b) to $[0, \infty]$:

- $D^+f(c) = \limsup_{x \to c^+} \dfrac{f(x) - f(c)}{x - c}$,
- $d^+f(c) = \liminf_{x \to c^+} \dfrac{f(x) - f(c)}{x - c}$,
- $D^-ff(c) = \limsup_{x \to c^-} \dfrac{f(x) - f(c)}{x - c}$,
- $d^-f(c) = \liminf_{x \to c^-} \dfrac{f(x) - f(c)}{x - c}$.

EXERCISE 5.10.9. Show that f is differentiable at c if and only if

$$D^+f(c) = d^+f(c) = D^-f(c) = d^-f(c)$$

and all four are finite.

Our strategy, therefore, will be to prove that the following sequence of inequalities holds for almost all $c \in (a, b)$:

$$D^+f(c) \leq d^-f(c) \leq D^-f(c) \leq d^+f(c) \leq D^+f(c).$$

EXERCISE 5.10.10. Show that two parts of this inequality are already clear from the definitions, namely,

$$d^-f(c) \leq D^-f(c) \quad \text{and} \quad d^+f(c) \leq D^+f(c).$$

It remains to be shown that

$$D^+f(c) \leq d^-f(c) \quad \text{and} \quad D^-f(c) \leq d^+f(c)$$

hold almost everywhere and $d^-f(x)$ and $D^+f(c)$ are finite almost everywhere.

EXERCISE 5.10.11. Show that it suffices to prove the first of these inequalities. (*Hint*: Consider $g(x) = -f(-x)$.)

Thus it suffices to show that for any monotone increasing function $f : [a, b] \longrightarrow \mathbb{R}$, we have that

$$D^+f(c) \leq d^-f(c)$$

for almost all $c \in (a, b)$ and that $D^+f(c)$ is finite almost everywhere.

The following exercise will be instrumental in the remainder of the argument.

EXERCISE 5.10.12. Let $g : (a, b) \longrightarrow \mathbb{R}$ be such that $g^-(x)$ and $g^+(x)$ exist for all $x \in (a, b)$, and set

$$G(x) = \max\{g^-(x), g(x), g^+(x)\}.$$

Then the set

$$S = \{x \in (a, b) \mid \text{there exists } y \in (x, b) \text{ with } g(y) > G(x)\}$$

is a finite or countable union of pairwise disjoint open intervals (a_k, b_k). Furthermore, for each such interval, $g^+(a_k) \leq G(b_k)$.

EXERCISE 5.10.13. Show that if $f : [a, b] \longrightarrow \mathbb{R}$ is monotone increasing, then the set of points where f is discontinuous is at most countable.

Since every countable subset of $\mathbb{R}$ has measure zero, this shows that the set of points where f is discontinuous has measure zero.

Define subsets $A, B \subseteq [a, b]$ by

$$A = \{c \in (a, b) \mid f \text{ is continuous at } c \text{ and } D^+f(c) = +\infty\}$$

and

$$B = \{c \in (a, b) \mid f \text{ is continuous at } c \text{ and } D^+f(c) > d^-f(c)\}.$$

EXERCISE 5.10.14. Show that A has measure zero by applying Exercise 5.10.12 to the functions $g(x) = f(x) - Nx$ for natural numbers $N \in \mathbb{N}$.

Define sets $B_{r,s}$ for $r, s \in \mathbb{Q}$ by

$$B_{r,s} = \{c \in (a, b) \mid f \text{ is continuous at } c,\ d^-f(c) < r, \text{ and } D^+f(c) > s\}.$$

EXERCISE 5.10.15. Show that $B_{r,s}$ has measure zero for all $r, s \in \mathbb{Q}$ as follows. Use Exercise 5.10.12 to construct inductively a sequence of sets S_i with the following properties.

- $B_{r,s}$ is contained in S_{2n} except for points that lie in intervals that were cut off by our procedure in the previous steps.
- The total length of the intervals in S_{2n} is at most $(\frac{r}{s})^n \cdot (b - a)$.
- The total length of the intervals that we have ignored is

$$\frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{8} + \cdots + \frac{\varepsilon}{2^n} < \varepsilon.$$

EXERCISE 5.10.16. Show that the set B has measure zero and finish the proof of Lebesgue's theorem.

5.10.2. The Riemann Integral in $\mathbb{R}^n$.

DEFINITION 5.10.17. Let $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ be points in $\mathbb{R}^n$ such that $a_i \leq b_i$ for all i. The *closed rectangle* $[a, b]$ is the set

$$\{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i \text{ for all } i\}.$$

The *open rectangle* (a, b) is the set

$$\{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i < x_i < b_i \text{ for all } i\}.$$

By a *rectangle* in $\mathbb{R}^n$ we mean the obvious variant of these in which endpoints may or may not be included (that is, the inequalities above are allowed to be strict or nonstrict).

The points a and b in the definitions above are sometimes referred to as the *endpoints* of the rectangle.

DEFINITION 5.10.18. If R is a rectangle with endpoints a and b, then the *volume* of R is

$$v(R) = \prod_{i=1}^{n}(b_i - a_i).$$

DEFINITION 5.10.19. Let A be a subset of $\mathbb{R}^n$. We say that A has *measure zero* if for every $\varepsilon > 0$, there exists a countable collection $\{R_k\}$ of open rectangles such that

$$A \subseteq \bigcup_{k=1}^{\infty} R_k$$

and

$$\sum_{k=1}^{\infty} v(R_k) \leq \varepsilon.$$

EXAMPLE 5.10.20.

(i) The empty set has measure zero.
(ii) Any countable set has measure zero.
(iii) Any subset of a set of measure zero has measure zero.

EXERCISE 5.10.21. Show that the Cantor set (see Exercise 1.6.26) has measure zero.

EXERCISE 5.10.22. Show that a subset A of $\mathbb{R}^n$ has measure zero if and only if there exists a countable collection $\{R_k\}$ of open rectangles such that

$$A \subseteq \bigcup_{k=1}^{\infty} R_k,$$

$$\sum_{k=1}^{\infty} v(R_k) < \infty,$$

and each point of A lies in an infinite number of the open rectangles R_k.

DEFINITION 5.10.23. Let $k \in \mathbb{Z}$, $1 \leq k \leq n$. If R is a rectangle, an $(n-k)$-dimensional face of R is a subset of R in which k of the inequalities are replaced by the equalities $x_i = a_i$ or $x_i = b_i$.

EXERCISE 5.10.24. Show that an $(n-k)$-dimensional face of a rectangle R has measure zero.

EXERCISE 5.10.25. Prove that the intersection of two rectangles is a rectangle.

DEFINITION 5.10.26. Let R be a closed rectangle. A *partition* of R is a finite collection of open rectangles $P = \{R_1, R_2, \ldots, R_m\}$ which are pairwise disjoint and satisfy

$$R = \bigcup_{i \in \mathbb{N}} \overline{R}_i.$$

EXERCISE 5.10.27. Prove the following.

(i) If $P = \{R_1, R_2\}$ is a partition of R, then $v(R_1) + v(R_2) = v(R)$.
(ii) If $P = \{R_1, R_2, \ldots, R_m\}$ is a partition of R ($m \geq 2$), there exists a partition $Q = \{S_1, S_2\}$ of R such that $R_1 \cap S_2 = R_2 \cap S_1 = \varnothing$.
(iii) Show by induction on m that if $P = \{R_1, \ldots, R_m\}$ is a partition of R, then $\sum_{i=1}^{m} v(R_i) = v(R)$.

EXERCISE 5.10.28. Let R be a closed rectangle and let S be an open rectangle with $S \subseteq R$. Show that there exists a partition P of R with $S \in P$.

If $R \subseteq \mathbb{R}^n$ is an arbitrary rectangle and $f : \overline{R} \longrightarrow \mathbb{R}$ is a bounded function, define

$$M_f(R) = \sup\{f(x) \mid x \in \overline{R}\} \quad \text{and} \quad m_f(R) = \inf\{f(x) \mid x \in \overline{R}\}.$$

EXERCISE 5.10.29. Show that if $S \subseteq R$ are open rectangles and f is bounded on $\overline{R}$, then f is bounded on S and $m_f(R) \leq m_f(S) \leq M_f(S) \leq M_f(R)$.

In the following discussion, R will denote a closed rectangle in $\mathbb{R}^n$ and $f : R \longrightarrow \mathbb{R}$ will denote a bounded function.

DEFINITION 5.10.30. If $P = \{R_1, \ldots, R_m\}$ is a partition of R, define

$$L(f, P) = \sum_{i=1}^{m} v(R_i) m_f(R_i)$$

and

$$U(f, P) = \sum_{i=1}^{m} v(R_i) M_f(R_i).$$

EXERCISE 5.10.31. Show that for any partition P of R, $L(f, P) \leq U(f, P)$.

DEFINITION 5.10.32. If P and P' are partitions of R, we say that P' is a *refinement* of P if for every $S \in P'$, $S \subseteq T$ for some $T \in P$.

EXERCISE 5.10.33. Show the following.

(i) If P' is a refinement of P, $L(f, P) \leq L(f, P') \leq U(f, P') \leq U(f, P)$.
(ii) If P_1 and P_2 are partitions of R, there exists a common refinement. That is, there exists a partition P which is a refinement of both P_1 and P_2.
(iii) If P_1 and P_2 are partitions of R, we have $L(f, P_1) \leq U(f, P_2)$.

DEFINITION 5.10.34. Let

$$\mathscr{L}(f, R) = \sup\{L(f, P) \mid P \text{ a partition of } R\}$$

and

$$\mathscr{U}(f, R) = \inf\{U(f, P) \mid P \text{ a partition of } R\}.$$

Note that by Exercise 5.10.33, $\mathscr{L}(f,R) \leq \mathscr{U}(f,R)$. We say f is *Riemann integrable* on R if $\mathscr{L}(f,R) = \mathscr{U}(f,R)$. In this case, we say the *Riemann integral* of f on R, written $\int_R f(x)dx$, equals $\mathscr{L}(f,R) = \mathscr{U}(f,R)$.

EXERCISE 5.10.35. Show that f is Riemann integrable on R iff for all $\varepsilon > 0$ there exists a partition P such that $U(f,P) - L(f,P) < \varepsilon$.

EXERCISE 5.10.36. Show that f is Riemann integrable on R iff there exists a sequence of partitions $P_1, P_2, \ldots, P_m, \ldots$ such that $\lim_{m\to\infty} U(f,P_m) = \lim_{m\to\infty} L(f,P_m)$, and in this case we have $\int_R f(x)dx$ equals this limit.

EXERCISE 5.10.37. Let $a, b \in \mathbb{R}$ with $a < b$. Using the result of Exercise 5.10.36, show that

(i) $f(x) = x$ is integrable on $[a,b]$ and $\int_{[a,b]} x dx = \frac{b^2-a^2}{2}$ and

(ii) $f(x) = x^k$ is integrable on $[a,b]$ and $\int_{[a,b]} x^k dx = \frac{b^{k+1}-a^{k+1}}{n+1}$, for all $k \in \mathbb{N}$.

DEFINITION 5.10.38. If $R = [a_1,b_1]\times[a_2,b_2]\times\cdots\times[a_n,b_n]$ is an arbitrary rectangle, define

$$d(R) = \sqrt{\sum_{j=1}^{n}(b_j - a_j)^2}.$$

This is the familiar diameter of R. In particular, for any $x, y \in R$, we have $\|x - y\| \leq d(R)$. If P is a partition of R, define $d(P) = \max\{d(R) \mid R \in P\}$.

EXERCISE 5.10.39. Show the following.

(i) For all $\delta > 0$, there exists a partition P of R such that $d(P) \leq \delta$.

(ii) If P' is a refinement of P, then $d(P') \leq d(P)$.

EXERCISE 5.10.40. Show that f is Riemann integrable on R iff for all $\varepsilon > 0$ there exists $\delta > 0$ such that $d(P) < \delta$ implies $U(f,P) - L(f,P) < \varepsilon$.

EXERCISE 5.10.41. Suppose that f and g are integrable of R and $c \in \mathbb{R}$. Then $cf + g$ is integrable on R and $\int_R (cf(x) + g(x))dx = c\int_R f(x)dx + \int_R g(x)dx$.

EXERCISE 5.10.42. Show that if f is continuous on the closed rectangle R, then f is Riemann integrable on R.

The following theorem, which we prove in exercises, characterizes Riemann integrable functions.

THEOREM 5.10.43. *The function f is Riemann integrable on R iff f is continuous almost everywhere on R.*

EXERCISE 5.10.44. If $R \subseteq \mathbb{R}^n$ is a rectangle, show that the boundary ∂R has measure zero.

We now prove, in an exercise, that if f is Riemann integrable on R, then f is continuous almost everywhere on R.

EXERCISE 5.10.45. Let K be the set of points in R where f is not continuous.

(i) For $N \in \mathbb{N}$, let $K_N = \{x \in R \mid$ for any open set U containing x there exists $y \in U \cap R$ such that $|f(y) - f(x)| > 1/N\}$. Explain why it is enough to show that each K_N has measure zero.

(ii) Show that if $S \subseteq R$ is an open rectangle with $K_N \cap S \neq \varnothing$, then $M_f(S) - m_f(S) > 1/N$.

(iii) Pick a partition P of R with $U(f,P) - L(f,P) < \varepsilon$. Find an upper bound for the expression

$$\sum_{\substack{S \in P \\ K_N \cap S \neq \varnothing}} v(S).$$

(iv) Using part (iii) and Exercise 5.10.44, argue that K_N has measure zero.

We now prove, in another exercise, that if f is continuous almost everywhere, then f is Riemann integrable on R.

EXERCISE 5.10.46. Assume that f is continuous almost everywhere on R.

(i) Let $C \subseteq R$ be the set of points at which f is continuous. Pick $\varepsilon > 0$. Then for each $x \in C$, there exists δ_x such that for $y \in R$ and $\|y - x\| < \delta_x$, $|f(y) - f(x)| < \varepsilon$. Let $U = \bigcup_{x \in C} B_{\delta_x/2}(x)$ and $V = R \setminus U$. Show that V is compact and has measure zero.

(ii) Show that there exist finitely many open rectangles $R_1, \ldots, R_m$ such that $\sum_{i=1}^m v(R_i) < \varepsilon$ and $V \subseteq \bigcup_{i=1}^m R_i$.

(iii) Let $W = R \setminus \bigcup_{i=1}^m R_i$. Argue that $W \subseteq U$ and W is compact. Show that there exists $\delta > 0$ such that for $x, y \in W$ and $\|x - y\| < \delta$, we have $|f(x) - f(y)| < 2\varepsilon$.

(iv) Using Exercises 5.10.28, 5.10.33, and 5.10.39, show that there exists a partition P of R with $d(P) < \delta$ such that for each $S \in P$, either $S \subseteq W$ or $S \subseteq R_i$ for some i.

(v) Prove that for such a partition P, we have $U(f,P) - L(f,P) < (M_f(R) - m_f(R))\varepsilon + 2\varepsilon v(R)$. Conclude that f is Riemann integrable.

We now extend the definition of the Riemann integral to arbitrary bounded sets in $\mathbb{R}^n$.

DEFINITION 5.10.47. Let $D \subseteq \mathbb{R}^n$ be a bounded subset and let $f : D \longrightarrow \mathbb{R}$ be a bounded function. If R is a closed rectangle containing D, we may extend f to a bounded function $\overline{f}$ on R by

$$\overline{f}(x) = \begin{cases} f(x), & x \in D, \\ 0, & x \in R \setminus D. \end{cases}$$

We say that f is Riemann integrable on D if there exists such a closed rectangle R such that $\overline{f}$ is Riemann integrable on R. In this case we define

$$\int_D f(x)dx = \int_R \overline{f}(x)dx.$$

EXERCISE 5.10.48. Show that this definition is independent of the choice of the rectangle R.

EXERCISE 5.10.49. Suppose that f and g are Riemann integrable on D and $f(x) \leq g(x)$ for all $x \in D$. Show that $\int_D f(x)dx \leq \int_D g(x)dx$.

EXERCISE 5.10.50. Suppose that f and g are Riemann integrable on D and that $c \in \mathbb{R}$. Show that $cf + g$ is Riemann integrable on D and

$$\int_D (cf(x) + g(x))dx = c\int_D f(x)dx + \int_D g(x)dx.$$

DEFINITION 5.10.51. A subset $D \subseteq \mathbb{R}^n$ is said to have *Jordan content zero* if, for every $\varepsilon > 0$, there exists a finite collection of rectangles covering D whose total volume is less than ε.

EXERCISE 5.10.52. Show that a compact subset $D \subseteq \mathbb{R}^n$ has Jordan content zero iff D has measure zero.

DEFINITION 5.10.53. A bounded subset $D \subseteq \mathbb{R}^n$ is a *Jordan domain* if its boundary has Jordan content zero.

EXERCISE 5.10.54. Let $D \subseteq \mathbb{R}^n$ be a Jordan domain, and let $f : D \longrightarrow \mathbb{R}$ be a bounded function. Show that if the set of discontinuities of f has measure zero, then f is Riemann integrable.

EXERCISE 5.10.55. Let $D \subseteq \mathbb{R}^n$ be a Jordan domain, and let $(f_i)_{i\in\mathbb{N}}$ be a sequence of Riemann integrable functions on D. Suppose that f_i converges uniformly to a function f. Show that f is Riemann integrable and that $\int_R f_i \to \int_R f$.

5.10.3. The Construction of C_c^∞ Functions in $\mathbb{R}^n$.

Let $C \subseteq \mathbb{R}^n$ be a compact set and let $U \subseteq \mathbb{R}^n$ be a bounded open set containing C. The goal of this project is to show that there is a nonnegative, compactly supported C^∞ function f such that $f = 1$ on C and f vanishes on the complement of U. Before we begin the discussion, you should note that this is similar to Urysohn's lemma (see the project in Section 2.7.1) in which the function f is required only to be continuous.

EXERCISE 5.10.56. Let

$$f(x) = \begin{cases} e^{-(x-1)^{-2}}e^{-(x+1)^{-2}} & \text{if } x \in (-1,1), \\ 0 & \text{if } x \notin (-1,1). \end{cases}$$

Show that $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a C^∞ function which is positive on $(-1,1)$ and 0 elsewhere.

EXERCISE 5.10.57. Given $k, a, b \in \mathbb{R}$ such that $a < b$, show that there is a C^∞ function $g : \mathbb{R} \longrightarrow [0,k]$ such that $g(x) = 0$ for $x \leq a$ and $g(x) = k$ for $x \geq b$.

EXERCISE 5.10.58. If $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, define $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ by

$$g(x) = f((x_1 - a_1)/\varepsilon) \cdots f((x_n - a_n)/\varepsilon).$$

Show that g is a C^∞ function which is positive on

$$(a_1 - \varepsilon, a_1 + \varepsilon) \times \cdots \times (a_n - \varepsilon, a_n + \varepsilon)$$

and 0 elsewhere.

EXERCISE 5.10.59. If $U \subseteq \mathbb{R}^n$ is bounded and open and $C \subseteq U$ is compact, show that there is a nonnegative C^∞ function $f : U \longrightarrow \mathbb{R}$ such that $f(x) > 0$ for $x \in C$ and $f = 0$ outside of some closed set contained in U.

EXERCISE 5.10.60. Show that we can choose such an f so that $f : U \longrightarrow [0, 1]$ and $f(x) = 1$ for $x \in C$.

Note that the results of this project will allow us to approximate functions in certain norms by C_c^∞ functions. However, it is not possible in general to approximate a continuous function with the supremum norm by a compactly supported function on all of $\mathbb{R}^n$.

Chapter 6

Integration

> Finally, the reader will probably observe the conspicuous absence of a time-honored topic in Calculus courses, the "Riemann integral". It may well be suspected that, had it not been for its prestigious name, this would have been dropped long ago, for (with due reverence to Riemann's genius) it is certainly quite clear to any working mathematician that nowadays such a "theory" has at best the importance of a mildly interesting exercise in the general theory of measure and integration. Only the stubborn conservatism of academic tradition could freeze it into a regular part of the curriculum, long after it had outlived its historical importance. Of course, it is perfectly feasible to limit the integration process to a category of functions which is large enough for all purposes of elementary Analysis (at the level of this course), but close enough to the continuous functions to dispense with any consideration drawn from measure theory; this is what we have done by defining only the integral of regulated functions (sometimes called the "Cauchy integral"). When one needs a more powerful tool, there is no point in stopping halfway, and the general theory of ("Lebesgue") integration is the only sensible answer.
>
> – *Jean Dieudonné*
> *Foundations of Modern Analysis*

Challenge Problems: VI

PROBLEM 6.1. Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function, and suppose that $f \circ u$ is C^∞ for every C^∞ function $u : \mathbb{R} \longrightarrow \mathbb{R}^n$. Show that f is C^∞.

PROBLEM 6.2. Show that if $r > 2$, then

$$f_r(x) = \begin{cases} 1/q^r & \text{if } x = p/q, \\ 0 & \text{if } x = 0 \text{ or if } x \text{ is irrational} \end{cases}$$

is differentiable almost everywhere.

PROBLEM 6.3. Show that if $r > 2$ and if α is a real algebraic number of degree greater than or equal to 2, then f_r is differentiable at α.

PROBLEM 6.4. Find $\inf\{r > 2 \mid f_r \text{ is differentiable at } \pi\}$. Do the same for e in place of π.

PROBLEM 6.5. Given a transcendental number $\alpha \in \mathbb{R}$, does there always exist $r > 2$ such that f_r is differentiable at α? In any case, describe those transcendental numbers $\alpha \in \mathbb{R}$ such that f_r is differentiable at α for some $r > 2$.

6.1. Measures

To develop a theory of integration in any space, we first need a notion of volume in that space. In the particular case of Euclidean space, the notion of volume comes from Lebesgue measure.

In this chapter, we will begin our discussion with a general theory of measure and then turn to a development of Lebesgue measure on Euclidean space. The general theory of measure and integration is really quite straightforward and involves a cast of characters that includes:

(1) σ-algebras of subsets of a given set,
(2) measures on σ-algebras,
(3) measurable functions, and
(4) integration of measurable functions.

DEFINITION 6.1.1. Let X be a set. A *σ-algebra on X* is a collection $\mathfrak{M}$ of subsets of X that contains the empty set and is closed under countable unions and complements.

Observe that there are two obvious σ-algebras on any set X. The first is the σ-algebra consisting of the empty set and X, and the second is the σ-algebra $\mathscr{P}(X)$ of all subsets of X. Neither of these examples will be of much interest to us, except in some special cases. For instance, when we consider $X = \mathbb{N}$ or $\mathbb{Z}$, we will probably want individual points to be members of our σ-algebra $\mathfrak{M}$. Hence, in these cases, $\mathfrak{M} = \mathscr{P}(\mathbb{N})$ or $\mathscr{P}(\mathbb{Z})$.

EXERCISE 6.1.2.

(i) Show that a σ-algebra is closed under set difference and countable intersection.
(ii) Show that a collection $\mathfrak{M}$ of subsets of X that contains the empty set and is closed under countable intersection and complementation is a σ-algebra.

Where do σ-algebras come from? One way to get a σ-algebra is to start with a nonempty collection S of subsets of X and then just take the intersection of all σ-algebras containing S. Note that the intersection of any collection of σ-algebras is again a σ-algebra. By our remark above, we know that $\mathscr{P}(X)$ is one such σ-algebra, so this actually means something.

DEFINITION 6.1.3. Let X be a set, and let S be a nonempty collection of subsets of X. The *σ-algebra $\mathfrak{M}(S)$ generated by S* is the intersection of all σ-algebras containing S.

EXERCISE 6.1.4.

(i) Take $X = \mathbb{R}$ and describe the σ-algebra generated by the collection of open intervals.

(ii) Let X be a separable metric space and let $\mathfrak{M}$ be the σ-algebra generated by open balls in X. Show that $\mathfrak{M}$ contains all the open sets in X and all the closed sets in X. Describe some sets in $\mathfrak{M}$ that are neither open nor closed. The σ-algebra $\mathfrak{M}$ is called the σ-algebra of *Borel sets* in X.

EXERCISE 6.1.5. Can a σ-algebra in a set X have cardinality $\aleph_0$?

The next definition contains a useful, and in fact necessary, idea.

DEFINITION 6.1.6. We adjoin a symbol ∞ to the nonnegative real numbers to create the *nonnegative extended real numbers* $[0, \infty]$. Arithmetic mostly extends in the expected way: $x + \infty = \infty$ for any $x \in [0, \infty]$, and $x \cdot \infty = \infty$ whenever $x > 0$. We make the convention that $0 \cdot \infty = 0$.

Now, we turn to the next member of our cast of characters, a measure on a σ-algebra.

DEFINITION 6.1.7. Let X be a nonempty set, and let $\mathfrak{M}$ be a σ-algebra on X. A *measure on* $(X, \mathfrak{M})$ is a function $\mu : \mathfrak{M} \longrightarrow [0, \infty]$ that is *countably additive*. That is, if $\{A_i\}_{i \in I}$ is a finite or countable family of pairwise disjoint members of $\mathfrak{M}$, then $\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$.

REMARK 6.1.8. There are two examples of totally uninteresting measures. One is the case when $A \in \mathfrak{M}$ implies $\mu(A) = \infty$, and the other is when $A \in \mathfrak{M}$ implies $\mu(A) = 0$. We can also have a mixture of these two. We reject these as unimportant, and consequently, we will assume that there is a set $A \in \mathfrak{M}$ such that $0 < \mu(A) < \infty$.

For simplicity of notation, we will often refer to a set X with a σ-algebra $\mathfrak{M}$ and a measure μ on $(X, \mathfrak{M})$ by writing (X, μ), as long as $\mathfrak{M}$ is clear from the context. In any case, the triple $(X, \mathfrak{M}, \mu)$, or the pair $(X, \mathfrak{M})$, or the pair (X, μ) is called a *measure space*.

EXERCISE 6.1.9. Show that, with the finiteness condition of Remark 6.1.8, $\mu(\varnothing) = 0$.

EXAMPLES 6.1.10.

(i) Let $X = \mathbb{N}$ or $\mathbb{Z}$, and let $\mathfrak{M} = \mathscr{P}(X)$. For any $A \subseteq X$, define $\mu(A) = \#(A)$, that is, the number of points in A. This is a measure on $(X, \mathfrak{M})$ called the *counting measure*.

(ii) Let X be a set, and let $\mathfrak{M}$ be a σ-algebra on X. Fix a point $x_0 \in X$. For $A \in \mathfrak{M}$, we define $\mu(A) = 1$ if $x_0 \in A$ and $\mu(A) = 0$ if $x_0 \notin A$. This defines a measure on $(X, \mathfrak{M})$, which is usually called *the delta measure at x_0* and is denoted by δ_{x_0}.

PROPOSITION 6.1.11. *Measures are monotonic; that is, if A, $B \in \mathfrak{M}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$. They are also continuous from above. This means that if $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing sequence of sets in $\mathfrak{M}$, then $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i)$. Moreover, if $A_1 \supseteq A_2 \supseteq \cdots$ is a decreasing sequence of sets in $\mathfrak{M}$ and if $\mu(A_k) < \infty$ for some k, then $\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i)$.*

Proof. Suppose first that A, $B \in \mathfrak{M}$ and $A \subseteq B$. Then $B = A \cup (B \setminus A)$. But $B \setminus A \in \mathfrak{M}$, so by additivity, $\mu(B) = \mu(A) + \mu(B \setminus A)$, and $\mu(B \setminus A) \geq 0$.

Next, if $A_i \in \mathfrak{M}$, $i \in \mathbb{N}$, and $A_1 \subseteq A_2 \subseteq \cdots$, let $B_i = A_{i+1} \setminus A_i$. Then the B_i are disjoint, so $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i) = \lim_{i\to\infty} \mu(A_i)$.

The rest of the proof is in the next exercise.

EXERCISE 6.1.12. Prove the statement about decreasing sequences in the previous proposition, and show that the condition that some $\mu(A_k) < \infty$ cannot be omitted.

REMARK 6.1.13. Nothing hard has happened yet. Why? Because the construction of interesting σ-algebras and measures is where the difficulty (and fun) lies. So now we get down to business with Lebesgue measure. The approach we take is through the definition and properties of outer measure. This approach applies not only to Lebesgue measure on $\mathbb{R}^n$, but to measures on any space.

The best kind of measures are σ-finite measures.

DEFINITION 6.1.14. A measure space is said to be *σ-finite* if it is the countable union of subsets of finite measure.

6.2. Lebesgue Measure

In Section 6.1, we outlined the framework for the general theory of measure. The principal focus in this book is Lebesgue measure on Euclidean space, and we now turn our attention in that direction. In the projects at the end of the chapter, we construct a measure on a p-adic field that is analogous to Lebesgue measure on $\mathbb{R}$, and we also construct Haar measure on a locally compact group. Throughout this section, X denotes a general measure space (X, μ), where μ is a measure on X, and, as noted above, the underlying σ-algebra is understood. Here, $\mathbb{R}^n$ denotes Euclidean space with the usual topology, and the letter m will be used exclusively to denote Lebesgue measure on $\mathbb{R}^n$. Some of the theorems we prove hold for a general measure space, and others are specifically for Lebesgue measure. The distinction will be clear from the context and notation.

Recall from Chapter 5 the definition of a rectangle.

DEFINITION 6.2.1. If $I_j = \{(a_j, b_j) \mid j = 1, \ldots, n\}$ is an n-tuple of intervals in $\mathbb{R}$, then $R = I_1 \times \cdots \times I_n$ is called an *open rectangle* in $\mathbb{R}^n$, and the *volume* of R is $m(R) = \prod_{j=1}^{n} (b_j - a_j)$. If, instead, we take $I_j = [a_j, b_j)$,

then the Cartesian product is called a *half-open rectangle* in $\mathbb{R}^n$ and its volume is defined in exactly the same way.

If $\{R_1, \ldots, R_k\}$ is a finite set of pairwise disjoint rectangles in $\mathbb{R}^n$, then $m\left(\bigcup_{j=1}^k R_j\right)$ is defined to be $\sum_{j=1}^k m(R_j)$. These definitions correspond to our usual understanding of volume in $\mathbb{R}^n$.

EXERCISE 6.2.2. Show that the σ-algebra generated by the collection of open rectangles is the same as the σ-algebra generated by the half-open rectangles.

DEFINITION 6.2.3. For any $A \subseteq \mathbb{R}^n$, define the *Lebesgue outer measure* $m^*(A)$ to be $\inf\{\sum_{j\in J} m(R_j)\}$, where the infimum is taken over all coverings $\{R_j\}_{j\in J}$ of A by at most countably many rectangles. It is perfectly acceptable for $m^*(A)$ to be ∞. In fact, it is unavoidable.

Observe that $m^*(A)$ is defined for *every* subset A of $\mathbb{R}^n$.

EXERCISE 6.2.4.

(i) If A is a rectangle (with or without faces included), show that $m^*(A) = m(A)$.
(ii) Show that the outer measure of a face of a rectangle is zero.

PROPOSITION 6.2.5. *The outer measure m^* is countably subadditive on $\mathscr{P}(\mathbb{R}^n)$. That is, if $\{A_j\}_{j\in J}$ is a finite or countable family of subsets of $\mathbb{R}^n$, not necessarily pairwise disjoint, then $m^*\left(\bigcup_{j\in J} A_j\right) \leq \sum_{j\in J} m^*(A_j)$.*

Proof. Pick $\varepsilon > 0$, and take a cover of each A_j by rectangles whose total volume does not exceed $m^*(A_j)+2^{-j}\varepsilon$. Then $m^*(\bigcup_{j\in J} A_j) \leq \sum_{j\in J} m^*(A_j)+\varepsilon$. Since ε is arbitrary, this proves the proposition.

PROPOSITION 6.2.6. *If A and B are subsets of $\mathbb{R}^n$ and $A \subseteq B$, then $m^*(A) \leq m^*(B)$. That is, m^* is* monotonic.

Proof. Easy.

We will also find use for the idea of *Lebesgue inner measure*.

DEFINITION 6.2.7. Suppose that A is a bounded subset of $\mathbb{R}^n$ and let R be a rectangle containing A. We define the *Lebesgue inner measure* of A to be

$$m_*(A) = m(R) - m^*(R \setminus A).$$

EXERCISE 6.2.8.

(i) Show that $m_*(A)$ is independent of the choice of the rectangle R as long as R contains A.
(ii) Show that $m_*(A) \leq m^*(A)$ for any bounded subset $A \subseteq \mathbb{R}^n$.
(iii) Extend the definition of Lebesgue inner measure to unbounded subsets of $\mathbb{R}^n$.

We will show later (Corollary 6.2.19) that a compact set K is "measurable" so that $m(K)$ is well-defined, and we will show that for a subset $A \subseteq \mathbb{R}^n$, we have

$$m_*(A) = \sup_{K \subseteq A} m(K)$$

where the supremum is taken over compact sets $K \subseteq A$.

DEFINITION 6.2.9. For any set X, we define an *outer measure* on X to be a map

$$\mu^* : \mathscr{P}(X) \longrightarrow [0, \infty]$$

such that μ^* is countably subadditive and nondecreasing and such that the outer measure of the empty set is 0.

Note that, in this definition, there is no bothersome σ-algebra to retard our progress.

There are several approaches to the definition of a Lebesgue measurable set in $\mathbb{R}^n$. We will encounter some of these either in the text or in exercises. Here, we take a direct approach using only outer measure.

DEFINITION 6.2.10. Let X be a set, and let μ^* be an outer measure on X. A subset $A \subseteq X$ is *measurable* with respect to the outer measure if, for every $E \subseteq X$,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A).$$

Denote the set of all measurable sets by $\mathfrak{M}(\mu^*)$. For $A \in \mathfrak{M}(\mu^*)$, we write $\mu(A) = \mu^*(A)$ and call this the *measure* of A.

In particular, subsets of $\mathbb{R}^n$ which are measurable with respect to Lebesgue outer measure m^* are called *Lebesgue measurable*, and if $A \in \mathfrak{M}(m^*)$, we call $m(A)$ the *Lebesgue measure* of A.

It is sometimes useful to write the definition of measurable set in the form

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap {}^cA).$$

You should stare at this definition for a while. It is not at all obvious that there are any measurable sets whatsoever aside from the empty set and X. Nonetheless, with this definition, we can start proving theorems about Lebesgue measurable sets right away. We first state a proposition that holds for any outer measure.

PROPOSITION 6.2.11. *If μ^* is an outer measure on a set X and if $A \subseteq X$ is such that $\mu^*(A) = 0$, then A is measurable and $\mu(A) = 0$.*

Proof. Let $E \subseteq X$ be any set. By subadditivity, $\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \setminus A)$. Since $\mu^*(A) = 0$ and μ^* is nondecreasing, we have $\mu^*(E \cap A) = 0$, so $\mu^*(E) \leq \mu^*(E \setminus A)$. Monotonicity furnishes the reverse inequality, so A is measurable by definition.

Fortunately, there are loads of Lebesgue measurable sets. We start with the most obvious necessary fact, that a rectangle, which is already "measurable" (since we have defined its measure), is measurable in the sense of the definition above.

PROPOSITION 6.2.12. *A rectangle in $\mathbb{R}^n$ is a Lebesgue measurable set.*

Proof. Let $A \subseteq \mathbb{R}^n$ be a rectangle, and let E be any subset of $\mathbb{R}^n$. If $m^*(E) = \infty$, then $m^*(E \setminus A) = \infty$, and that does it. Otherwise, given $\varepsilon > 0$, cover E with a countable family of rectangles $\{R_j\}_{j\in J}$ such that $\sum_{j\in J} m(R_j) < m^*(E) + \varepsilon$. Then, for each rectangle R_j in this cover, $R_j \cap A$ is itself a rectangle. Also, $R_j \setminus A$ can be written as a finite union of rectangles. So now we have subdivided the rectangles in the original cover into a new countable cover $\{S_j\}_{j\in J'}$, and $\sum_{j\in J'} m(S_j) = \sum_{j\in J} m(R_j)$. Finally, those S_j that intersect A (and hence are contained in it) cover $E \cap A$, whereas those that do not intersect A cover $E \setminus A$. Consequently, $m^*(E\cap A) + m^*(E \setminus A) < m^*(E) + \varepsilon$.

REMARK 6.2.13. We originally defined the volume of a rectangle in the usual fashion. We then observed that $m^*(R) = m(R)$ for any rectangle R. Now that we have proved that rectangles are measurable in the general sense, we can write $m(R)$ without any ambiguity.

PROPOSITION 6.2.14. *A finite union of rectangles is Lebesgue measurable.*

Proof. Same proof with more dissections.

Obviously the next step is to prove that a countable union of rectangles is measurable. We might be tempted to try the methods used above, but this just leads to frustration. We cut right to the chase and prove that the measurable sets relative to *any* outer measure form a σ-algebra.

THEOREM 6.2.15. *Let X be a set and let μ^* be an outer measure on X. Then, the measurable sets with respect to μ^* form a σ-algebra, and the restriction of μ^* to this σ-algebra is a measure.*

Proof. First we need to show that the union of two measurable sets is measurable. If A, B are measurable sets and $E \subseteq X$ is any set, then

$$\begin{aligned} \mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \setminus A) \\ &= \mu^*(E \cap A) + \mu^*((E \setminus A) \cap B) + \mu^*((E \setminus A) \setminus B) \\ &\geq \mu^*(E \cap (A \cup B)) + \mu^*(E \setminus (A \cup B)). \end{aligned}$$

Thus $A \cup B$ is measurable (the opposite inequality is forced by the definition of outer measure). Also, the complement of a measurable set is measurable, since the equation $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap {}^cA)$ is really symmetric in A and ${}^cA$.

Now, if $\{A_i\}_{i\in I}$ is a finite or countable family of disjoint measurable sets, we need to show that $\mu^*\left(\bigcup_{i\in I} A_i\right) \geq \sum_{i\in I} \mu^*(A_i)$; again, we already

have the opposite inequality. Set $B_j = \bigcup_{i=1}^{j} A_i$. The sets B_j are measurable since each A_i is measurable. We have

$$\mu^*(B_{j+1}) = \mu^*(B_{j+1} \cap A_{j+1}) + \mu^*(B_{j+1} \setminus A_{j+1}) = \mu^*(A_{j+1}) + \mu^*(B_j).$$

By induction, $\mu^*(B_j) = \sum_{i=1}^{j} \mu^*(A_i)$. But $\mu^* \left(\bigcup_{i\in I} A_i\right) \geq \mu^*(B_j)$ for every j, which proves that μ^* is countably additive on the measurable sets A_i.

Finally, we show that if $\{A_i\}_{i\in I}$ is a countable family of measurable sets, then $\bigcup_{i\in I} A_i$ is measurable. Again, we set $B_j = \bigcup_{i=1}^{j} A_i$ and $A = \bigcup_{i\in I} A_i = \bigcup_{j\in I} B_j$.

If $E \subseteq X$, we define a new outer measure $\mu^*|_E$ on X by $\mu^*|_E(A) = \mu^*(E \cap A)$.

Exercise 6.2.16. Show that $\mu^*|_E$ is an outer measure. (As the notation suggests, $\mu^*|_E$ is called the *restriction* of μ^* to E.) Show that any set that is measurable with respect to μ^* is measurable with respect to $\mu^*|_E$ (regardless of whether E is measurable).

Now if $\mu^*(E) = \infty$, the inequality $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \setminus A)$ is automatic. So assume $\mu^*(E) < \infty$. Since the B_j form an increasing sequence of $\mu^*|_E$-measurable sets, we can write

$$\begin{aligned}
\mu^*(E \cap A) + \mu^*(E \setminus A) &= \mu^*|_E(A) + \mu^*|_E({}^cA) \\
&= \mu^*|_E\left(\bigcup_{j\in I} B_j\right) + \mu^*|_E\left({}^c\bigcup_{j\in I} B_j\right) \\
&= \lim_{j\to\infty} \mu^*|_E(B_j) + \lim_{j\to\infty} \mu^*|_E({}^cB_j) \\
&= \mu^*(E).
\end{aligned}$$

The second-to-last equality follows from a slight modification of Proposition 6.1.11, since we have already shown that $\mu^*|_E$ is countably additive. Thus, A is μ^*-measurable.

This powerful theorem has some immediate consequences.

Corollary 6.2.17. *A countable union of rectangles is Lebesgue measurable.*

Corollary 6.2.18. *Open sets in $\mathbb{R}^n$ are Lebesgue measurable.*

Proof. Open rectangles with rational vertices form a countable basis for the topology on $\mathbb{R}^n$.

Corollary 6.2.19. *Compact sets in $\mathbb{R}^n$ are Lebesgue measurable, and if K is compact, then $m(K) < \infty$.*

Exercise 6.2.20. Show that a bounded set $A \subseteq \mathbb{R}^n$ is Lebesgue measurable if and only if $m_*(A) = m^*(A)$.

We now consider the σ-algebra of subsets of $\mathbb{R}^n$ generated by the open sets. The members of this σ-algebra are called the *Borel sets*.

COROLLARY 6.2.21. *Borel sets in $\mathbb{R}^n$ are Lebesgue measurable.*

Proof. The σ-algebra generated by the open sets is a subalgebra of the σ-algebra of all Lebesgue measurable sets, by Corollary 6.2.18.

REMARK 6.2.22. You have to be very careful here. Borel sets are certainly the most important among the Lebesgue measurable sets in $\mathbb{R}^n$, but there are many more Lebesgue measurable sets than Borel sets. In particular, if A is a Borel set and B is a set of Lebesgue measure zero, then $A \cup B$ is Lebesgue measurable but need not be a Borel set. This will be made more specific later on.

EXERCISE 6.2.23. Prove that the cardinality of the Borel sets in $\mathbb{R}^n$ is c.

Here is a nice concrete geometric description of Lebesgue measurable sets.

PROPOSITION 6.2.24. *Suppose A is a bounded subset of $\mathbb{R}^n$. Then A is Lebesgue measurable if and only if, for any $\varepsilon > 0$, there is a finite union P of rectangles such that $m^*(A \triangle P) < \varepsilon$.*

Proof. Suppose A is measurable. Then, cover A with a countable family of rectangles $\{R_i\}_{i=1}^{\infty}$ such that $\sum_{i=1}^{\infty} m(R_i) < m(A) + \varepsilon$. Choose $N \in \mathbb{N}$ such that $\sum_{i=1}^{N} m(R_i) > \sum_{i=1}^{\infty} m(R_i) - \varepsilon$. Let $P = \bigcup_{i=1}^{N} R_i$. Then, $A \setminus P \subseteq \bigcup_{i=N+1}^{\infty} R_i$, so $m^*(A \setminus P) \leq \sum_{i=N+1}^{\infty} m(R_i) < \varepsilon$. Since A is Lebesgue measurable, $m^*(P \setminus A) = m^*(P) - m^*(P \cap A)$. Since P is Lebesgue measurable, $m^*(A \setminus P) = m(A) - m^*(P \cap A)$. We conclude that $m^*(P \setminus A) = m^*(A \setminus P) + m^*(P) - m(A) < \varepsilon + m^*(P) - m(A)$. But, certainly, $m^*(P) < \sum_{i=1}^{\infty} m(R_i) < m(A) + \varepsilon$. So $m^*(P \setminus A) < 2\varepsilon$.

In the other half of the proof, we get to practice our set theory. Suppose that, given $\varepsilon > 0$, we can always find a finite union P of rectangles such that $m^*(A \triangle P) < \varepsilon$. Then, for a fixed $E \subseteq \mathbb{R}^n$, we have

$$\begin{aligned}
&m^*(E \cap A) + m^*(E \setminus A) \\
&\quad = m^*(E \cap A \cap P) + m^*((E \cap A) \setminus P) + m^*((E \setminus A) \cap P) \\
&\qquad + m^*((E \setminus A) \setminus P) \\
&\quad = m^*(E \cap (A \cap P)) + m^*(E \setminus (A \cup P)) \\
&\qquad + m^*(E \cap (A \setminus P)) + m^*(E \cap (P \setminus A)) \\
&\quad \leq m^*(E \cap P) + m^*(E \setminus P) + 2m^*(E \cap (A \triangle P)) \\
&\quad \leq m^*(E) + 2\varepsilon.
\end{aligned}$$

Since ε was arbitrary, $m^*(E \cap A) + m^*(E \setminus A) \leq m^*(E)$, and A is measurable.

We now consider Borel measures on a metric space X. The σ-algebra of Borel sets on X is the σ-algebra generated by the open sets in X.

DEFINITION 6.2.25. A measure μ on a metric space X is a *Borel measure* if all Borel sets are μ-measurable. A Borel measure μ is *outer regular* if, for every measurable set A,

$$\mu(A) = \inf\{\mu(V) \mid V \text{ is open and } A \subseteq V\}.$$

A Borel measure μ is *inner regular* if, for every measurable set A,

$$\mu(A) = \sup\{\mu(K) \mid K \text{ is compact and } K \subseteq A\}.$$

An outer and inner regular Borel measure on a metric space X is called a *regular measure*.

EXERCISE 6.2.26.

(i) Show that the counting measure on $\mathbb{Z}$ with the discrete metric is regular.
(ii) Show that the delta measure with respect to a point x_0 on any metric space X is regular.

DEFINITION 6.2.27. A measure μ on a space X is called *complete* if, for any measurable set $A \subseteq X$ such that $\mu(A) = 0$, any subset $B \subseteq A$ is measurable.

THEOREM 6.2.28. *Lebesgue measure on $\mathbb{R}^n$ is regular and complete.*

Proof. Outer regularity is built into the construction of Lebesgue measure. Inner regularity is a bit more complicated. Fix a Lebesgue measurable set A with $m(A) < \infty$. Take $\varepsilon > 0$, and choose a large closed cube Q such that $m(A \cap Q) > m(A) - \varepsilon$ (if A is bounded, Q may well contain all of A). The set $Q \setminus A$ is also measurable, so choose an open set $V \supseteq Q \setminus A$ such that $m(V) < m(Q \setminus A) + \varepsilon$. Then, $Q \setminus V$ is a compact subset of A (since it is closed and bounded) and $m(Q \setminus V) > m(Q \cap A) - \varepsilon > m(A) - 2\varepsilon$.

To see that Lebesgue measure is complete, we simply refer to Proposition 6.2.11.

EXERCISE 6.2.29.

(i) Show that Lebesgue measure is σ-finite.
(ii) Show that the proof of inner regularity can be extended to Lebesgue measurable sets of infinite measure.

In the next proposition, we characterize Borel sets amongst Lebesgue measurable sets in $\mathbb{R}^n$. This was promised earlier.

PROPOSITION 6.2.30. *Let $A \subseteq \mathbb{R}^n$ be Lebesgue measurable. Then there exists a Lebesgue measurable set B having Lebesgue measure zero such that $A \setminus B$ is a Borel set.*

Proof. If A is bounded, approximate A from below by compact sets whose measures approach $m(A)$. The union of these sets is a Borel set C and $m(A \setminus C) = 0$. If A is unbounded, just divide $\mathbb{R}^n$ into countably many cubes and perform the same procedure in each cube.

At this point, we want to consider what happens to a Lebesgue measurable set under the action of a continuous map from $\mathbb{R}^n$ to $\mathbb{R}^n$. We will have more to say about this in the next section where we talk about measurable functions. For the moment, we ask the question, "If $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a homeomorphism and A is a Lebesgue measurable set in $\mathbb{R}^n$, under what conditions on f is $f(A)$ a Lebesgue measurable set?"

THEOREM 6.2.31. *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a homeomorphism. If f takes sets of Lebesgue measure zero to sets of Lebesgue measure zero, then f takes Lebesgue measurable sets to Lebesgue measurable sets.*

Proof. Let A be a Lebesgue measurable set. Then A can be written as the union of a Borel set B and a set N of measure zero. Since a homeomorphism preserves open and closed sets, it is clear that $f(B)$ is a Borel set. By assumption, $f(N)$ is measurable and has measure zero. Consequently, $f(A) = f(B) \cup f(N)$ is measurable.

REMARK 6.2.32. It is not generally true that the homeomorphic image of a set of measure zero is a set of measure zero. We shall discuss an example below. However, we do have the following general theorem.

THEOREM 6.2.33. *Suppose $A \subseteq \mathbb{R}^n$ and suppose $f : A \longrightarrow \mathbb{R}^n$ is a function such that for all $x \in A$,*

$$\limsup_{\substack{y \to x \\ y \in A}} \frac{\|f(x) - f(y)\|}{\|x - y\|} < \infty.$$

If $m(A) = 0$, then $m(f(A)) = 0$.

Proof. For integers $i, j \geq 1$, let

$$A_{ij} = \{x \in A \mid |f(x) - f(y)| \leq i|x - y| \text{ for all } y \in B_{1/j}(x) \cap A\}.$$

We first show that $A = \bigcup_{i,j} A_{ij}$. Let $x \in A$. Since $\limsup_{y \to x} \frac{\|f(x)-f(y)\|}{\|x-y\|} = L < \infty$, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $\frac{\|f(x)-f(y)\|}{\|x-y\|} < L + \varepsilon$ for all $y \in B_\delta(x) \cap A$. Choose $\varepsilon > 0$ such that $L + \varepsilon = i$ is a positive integer, let δ be as above, and choose a positive integer j such that $1/j < \delta$. It follows immediately that $x \in A_{ij}$.

Now fix i and j. Next we show that $m^*(f(A_{ij})) = 0$. Since $A_{ij} \subseteq A$, $m^*(A_{ij}) = 0$. Thus, for any $\varepsilon > 0$ there exist a sequence $(x_k)_{k \in \mathbb{N}}$ of elements of A and real numbers r_k with $0 < r_k < 1/j$ such that $A_{ij} \subseteq \bigcup_k B_{r_k}(x_k)$ and $\sum_{k \in \mathbb{N}} m(B_{r_k}(x_k)) < \varepsilon$. If $x \in A_{ij} \cap B_{r_k}(x_k)$, then $\|x - x_k\| < 1/j$, so $\|f(x) - f(x_k)\| \leq i\|x - x_k\| < ir_k$, and so we see that

$$f(A_{ij} \cap B_{r_k}(x_k)) \subseteq B_{ir_k}(f(x_k)).$$

Thus

$$f(A_{ij}) = \bigcup_k f(A_{ij} \cap B_{r_k}(x_k)) \subseteq \bigcup_k B_{ir_k}(f(x_k)),$$

and hence

$$m^*(f(A_{ij})) \leq \sum_k m^*(B_{ir_k}(f(x_k))) = i^n \sum_k m^*(B_{r_k}(x_k)) < i^n \varepsilon.$$

As this is true of all $\varepsilon > 0$, we have that $m^*(f(A_{ij})) = 0$, so

$$m^*(f(A)) = m^* \left(\bigcup_{i,j} f(A_{ij}) \right) \leq \sum_{i,j} m^*(f(A_{ij})) = 0.$$

By Proposition 6.2.11, $f(A)$ is Lebesgue measurable and has Lebesgue measure zero.

Exercise 6.2.34.

(i) Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a homeomorphism which is also absolutely continuous. Show that f carries sets of Lebesgue measure zero to sets of Lebesgue measure zero.

(ii) Let $c : [0,1] \longrightarrow \mathbb{R}$ be the Cantor function constructed in Example 2.5.6. Extend c to all of $\mathbb{R}$ by making $c(x) = 0$ for $x \leq 0$ and $c(x) = 1$ if $x \geq 1$. Now let $f(x) = c(x) + x$. Show that f is a homeomorphism of $\mathbb{R}$ that carries a set of measure zero (the Cantor set) to a set of positive measure.

(iii) Define $g : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ by $g(x,y) = (f(x), f(y))$, where f is the function defined in (ii). Show that g is a homeomorphism of $\mathbb{R}^2$ that takes a set of measure zero to a set of positive measure.

Corollary 6.2.35. *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a function, and assume that f satisfies a* Lipschitz condition of order 1. *That is, assume that there exists a constant $M \geq 0$ such that for all $x, y \in \mathbb{R}^n$,*

$$\|f(x) - f(y)\| \leq M\|x - y\|.$$

Then f carries sets of measure zero to sets of measure zero.

Corollary 6.2.36. *Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a nonsingular linear transformation. If $A \subseteq \mathbb{R}^n$ with $m(A) = 0$, then $m(T(A)) = 0$.*

Proof. It is clear that T satisfies the hypotheses of Corollary 6.2.35 with $M = \|T\|$.

Corollary 6.2.37. *Suppose that A is a Lebesgue measurable subset of $\mathbb{R}^n$ and that $T \in GL_n(\mathbb{R})$. Then $T(A)$ is measurable.*

The question is, "How does the action of a linear transformation alter the measure of a measurable set?"

Theorem 6.2.38. *Let $D = \operatorname{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$, that is, $D(x) = D(x_1, x_2, \ldots, x_n) = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$. If A is a measurable set, then $D(A) = \{Dx \mid x \in A\}$ is measurable and $m(D(A)) = |\det(D)| m(A)$.*

Proof. Notice that the transformation D preserves the class of rectangles in $\mathbb{R}^n$. Since the statement is immediate for any such rectangle, the result for measurable sets follows.

LEMMA 6.2.39. *Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be a translation, that is, $T(x) = x + x_0$ for some fixed $x_0 \in \mathbb{R}^n$. Then if A is measurable, $T(A)$ is measurable, and $m(T(A)) = m(A)$.*

Proof. Obvious.

THEOREM 6.2.40. *Suppose that $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is an isometry. If $A \subseteq \mathbb{R}^n$ is Lebesgue measurable, then $m(T(A)) = m(A)$.*

Proof. By translating if necessary, we can assume that $T \in O_n(\mathbb{R})$ (Mazur-Ulam and Exercise B.6.47). It is enough to show that the theorem holds for one reflection, for example, $R = \text{diag}(-1, 1, \ldots, 1)$, and all rotations. The former follows from Theorem 6.2.38, so we will focus on rotations. The remainder of the proof depends on the following lemma.

LEMMA 6.2.41. *Let R be a half-open rectangle in $\mathbb{R}^n$ and let U be any open set in $\mathbb{R}^n$. Then U can be written as the disjoint union of half-open rectangles similar to R.*

Proof. Without loss of generality we may take R to be a rectangle of the form

$$R = [0, a_1) \times \cdots \times [0, a_n)$$

for some positive real numbers $a_1, \ldots, a_n$. The obvious collection of translates of R is a cover $\mathcal{C}_0$ of $\mathbb{R}^n$ by disjoint half-open rectangles. Now for each integer $m \geq 0$, consider the rectangle $2^{-m}R$ and the analogous cover $\mathcal{C}_m$ of $\mathbb{R}^n$ by disjoint half-open translates of $2^{-m}R$. This sequence of covers $\mathcal{C}_0, \mathcal{C}_1, \ldots$ has the property that each cover is *refined* by all those that follow it. That is, for any $m' \geq m$, each rectangle in $\mathcal{C}_m$ is the disjoint union of those rectangles in $\mathcal{C}_{m'}$ that meet it.

Let $U \subseteq \mathbb{R}^n$ be any open set, and let

$$\mathcal{C}(U)_m = \{S \in \mathcal{C}_m \mid S \subseteq U\}$$

be the collection of rectangles in the cover $\mathcal{C}_m$ that are contained in U. The rectangles in $\bigcup_{m \geq 0} \mathcal{C}(U)_m$ are not necessarily disjoint, but using the above refinement property, we can eliminate redundancy in the collection in order to arrive at a disjoint collection as follows.

First, let $\mathcal{C}(U)'_0 = \mathcal{C}(U)_0$. Now take $\mathcal{C}(U)'_1$ to be the collection of rectangles in $\mathcal{C}(U)_1$ that do not meet any rectangles in $\mathcal{C}(U)'_0$. Then the union $\mathcal{C}(U)'_0 \cup \mathcal{C}(U)'_1$ is still a collection of disjoint rectangles. Moreover, the union of these rectangles is equal to the union of the rectangles in $\mathcal{C}(U)_1$ by the refinement property of the covers $\mathcal{C}_m$. Now proceed in the same fashion inductively by setting $\mathcal{C}(U)'_{m+1}$ equal to the collection of rectangles in $\mathcal{C}(U)_{m+1}$ that do not meet any rectangles in $\mathcal{C}(U)_m$. At each stage we see that the set

$$\mathcal{C}(U)'_0 \cup \mathcal{C}(U)'_1 \cup \cdots \cup \mathcal{C}(U)'_m$$

consists of disjoint rectangles and that the union of these rectangles is equal to the union of the rectangles in $\mathcal{C}(U)_m$.

Let $\mathcal{C}' = \bigcup_{m=0}^{\infty} \mathcal{C}(U)'_m$. This is a collection of disjoint rectangles similar to R, and we claim that the union of these rectangles is U. Indeed, this union coincides with the union of all of the rectangles in $\mathcal{C}(U)_m$ for all m by the above. We claim that the latter union is equal to U. Take any $x \in U$. Since U is open, there is a small open ball B containing x that is contained in U. For each m, there is exactly one element $S_m \in \mathcal{C}_m$ containing x. For m sufficiently large, S_m must be contained in B by diameter considerations, and hence contained in U, so $S_m \in \mathcal{C}(U)_m$.

We remark that this lemma is false if half-open rectangles are replaced by open rectangles.

Continuation of proof of Theorem 6.2.40. Let T be a rotation. From the construction of Lebesgue measure, it is enough to prove that the measure of any rectangle R is invariant under the action of T. We may take R to be a half-open rectangle since faces (and therefore their images under linear transformations) have measure zero. Let R' be any half-open rectangle similar to R, so that $R' = x + a \cdot R$ for some $x \in \mathbb{R}^n$ and $a > 0$. By translation invariance and Theorem 6.2.38, we have $m(R') = a^n m(R)$ and $m(T(R')) = a^n m(T(R))$. It follows that the number $C = m(T(R'))/m(R')$ is independent of R' and depends only on R.

By the above lemma, the open unit ball B in $\mathbb{R}^n$ can be written as the countable disjoint union of a family of half-open rectangles $\{R_i\}_{i \in I}$ similar to R. Since $T(B) = B$, we have

$$\sum_{i \in I} m(R_i) = \sum_{i \in I} m(T(R_i)) = C \sum_{i \in I} m(R_i).$$

Since $\sum_{i \in I} m(R_i)$ is positive and finite, $C = 1$.

We have now determined the effects of orthogonal and diagonal matrices on the measures of Lebesgue measurable subsets of $\mathbb{R}^n$. It remains to determine the action of unipotent matrices, that is, the elements of N as defined in the project in Section B.6.1. In fact, we will show that elements of N preserve Lebesgue measure using Fubini's theorem in Section 6.6. It will then follow from the Iwasawa decomposition (Challenge Problem 5.5) that if $T \in GL_n(\mathbb{R})$ and A is a measurable subset of $\mathbb{R}^n$, then $m(T(A)) = |\det T| \cdot m(A)$.

The reader might wonder at this point whether all sets are measurable and if outer measure is just a subterfuge to make the mathematics look more difficult. Absolutely not! In fact, with Lebesgue measure on $\mathbb{R}^n$, nonmeasurable sets abound. This is the content of the following proposition.

PROPOSITION 6.2.42. *If $A \subseteq \mathbb{R}^n$ is a Lebesgue measurable set and $0 < m(A) < \infty$, then A has a nonmeasurable subset.*

Proof. Suppose every subset of A is measurable. Let T be a subset of $\mathbb{R}^n$ that contains one element from every coset of the additive subgroup $\mathbb{Q}^n$ in $\mathbb{R}^n$, so that $\mathbb{R}^n = \bigcup_{t\in T}(t + \mathbb{Q}^n) = \bigcup_{r\in\mathbb{Q}^n}(T + r)$. Write $A = \bigcup_{r\in\mathbb{Q}^n} A_r$, where $A_r = A \cap (T + r)$. By hypothesis, each A_r is measurable. We will show that $m(A_r) = 0$, which will force $m(A) = 0$.

Let $Q = \mathbb{Q}^n \cap [0, 1]^n$ be the set of rational points in the unit cube. If $K \subseteq A_r$ is compact, let $H = \bigcup_{q\in Q}(K + q)$. Then H is bounded, so $m(H) < \infty$; but this forces $m(K+q) = m(K) = 0$, since the translates $K+q$ are pairwise disjoint and have the same measure by translation invariance. Since every compact subset of A_r has measure zero, so does A_r, by regularity. Since A is a countable union of the A_r, A must have measure zero.

6.3. Measurable Functions

Well, now that we have these measures, what are we going to do with them? The process continues as follows. We define functions, called measurable functions, and after we study these a bit, we integrate them. Then we are really in business.

Definition 6.3.1. Let μ be a measure on a set X. A function $f : X \longrightarrow [-\infty, \infty]$ is *measurable* if, for any $a \in \mathbb{R}$, the set $f^{-1}(a, \infty]$ is measurable in X.

Remark 6.3.2. Actually, we do not need a measure on X in order to define a measurable function. What we need is the concept of a measurable set. In our development, this is furnished by the outer measure. When developing measures through σ-algebras, the definition of measurable function would appear as follows. Let X be a set, and let $\mathfrak{M}$ be a σ-algebra on X. Then $f : X \longrightarrow [-\infty, \infty]$ is a measurable function if, for any real number $a \in \mathbb{R}$, we have $f^{-1}(a, \infty] \in \mathfrak{M}$.

Exercise 6.3.3. If $f : X \longrightarrow [-\infty, \infty]$ is measurable and $B \subseteq \mathbb{R}$ is any Borel set, show that $f^{-1}(B)$ is measurable in X.

The first thing to observe here is that this is analogous to the definition of a continuous map if X is a metric space. That is, $f : X \longrightarrow \mathbb{R}$ is continuous if, for any open set $U \subseteq \mathbb{R}$, $f^{-1}(U)$ is open in X. Of course, in the above definition, X is just a measure space, not necessarily a metric space, so continuity is not a meaningful concept in this context. However, if X is a metric space as well as a measure space, there are some meaningful statements one can make about the relation between continuity and measurability.

Exercise 6.3.4. Suppose X is a metric space as well as a measure space, and suppose μ is a Borel measure on X. Show that a continuous function $f : X \longrightarrow \mathbb{R}$ is measurable.

Measurability is a very nice property for a function to have. In fact, it is absolutely essential if we wish to integrate it. However, when pushing sets

forward or pulling them back, we have to be very careful about the class of sets on each end.

EXERCISE 6.3.5. Suppose X and Y are measure spaces. Define $f : X \longrightarrow Y$ to be measurable if for any measurable set $B \subseteq Y$, $f^{-1}(B)$ is measurable in X. Does this definition of measurability cause problems? *Hint*: Think about Exercise 6.2.34 and Proposition 6.2.42.

EXERCISE 6.3.6 **(Justice being served)**. Prove that the generalized Cantor set C_α constructed in Chapter 3 is Lebesgue measurable and has measure $1 - \alpha$.

EXERCISE 6.3.7. Suppose $f : X \longrightarrow [-\infty, \infty]$ is a function. Show that f is measurable iff $f^{-1}(a, \infty]$ is measurable for all $a \in \mathbb{Q}$. Indeed, for extended real-valued functions, show that we can select from the following menu: The function f is measurable iff one of the following is true for all rational a:

(i) $f^{-1}(a, \infty]$ is measurable;
(ii) $f^{-1}[a, \infty]$ is measurable;
(iii) $f^{-1}[-\infty, a)$ is measurable;
(iv) $f^{-1}[-\infty, a]$ is measurable.

EXERCISE 6.3.8.

(i) Suppose X is a measure space, $f : X \longrightarrow \mathbb{R}$ is measurable, and $g : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous. Show that $g \circ f : X \longrightarrow \mathbb{R}$ is measurable.
(ii) Find an example of a measure space X and measurable functions $f : X \longrightarrow \mathbb{R}$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$ such that $g \circ f$ is not measurable.
(iii) Suppose X is a measure space, $f : X \longrightarrow \mathbb{R}$ is a measurable function, and $g : \mathbb{R} \longrightarrow \mathbb{R}$ is such that $g^{-1}(V)$ is a Borel set for all open $V \subseteq \mathbb{R}$. Show that $g \circ f$ is a measurable function.

PROPOSITION 6.3.9. *The class of measurable functions $f : X \longrightarrow [-\infty, \infty]$ is closed under sums, products, countable supremum, countable infimum,* $\limsup$, *and* $\liminf$. *Hence, this class is a lattice, which is closed under pointwise limits.*

Proof. What we do here is to prove one of these statements to help the reader along, and we leave the rest to the reader to impede their progress. Let $f, g : X \longrightarrow [-\infty, \infty]$ be measurable functions, and consider $f \vee g$. For $a \in \mathbb{R}$, $(f \vee g)^{-1}(a, \infty] = f^{-1}(a, \infty] \cup g^{-1}(a, \infty]$, which is measurable.

The reader should note that the class of measurable functions is one of the things for which we have been searching for some time, that is, a reasonable class of functions that is closed under pointwise limits. We also have the following fact that is useful in integration theory.

EXERCISE 6.3.10. If $f : X \longrightarrow [-\infty, \infty]$ is a function, define functions f^+ and f^- on X by

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0, \\ 0 & \text{if } f(x) < 0 \end{cases} \quad \text{and} \quad f^-(x) = \begin{cases} 0 & \text{if } f(x) > 0, \\ -f(x) & \text{if } f(x) \leq 0. \end{cases}$$

Show that $f = f^+ - f^-$ and that if f is measurable, then so are f^+ and f^-.

DEFINITION 6.3.11. If something happens on a measure space except on a set of measure zero, we say that it happens *almost everywhere*. "Almost everywhere" is often abbreviated "a.e.".

For example, a sequence of functions $(f_i : X \longrightarrow \mathbb{R})_{i\in\mathbb{N}}$ converges pointwise almost everywhere if the sequence $(f_i(x))_{i\in\mathbb{N}}$ converges for almost every $x \in X$, that is, for every x except for members of a set of measure zero. In the theory of measure and integration that we are constructing, sets of measure zero are almost always irrelevant. Nonetheless, to be absolutely precise, we will always use the phrase *almost everywhere* when appropriate.

EXERCISE 6.3.12. Suppose $(f_i)_{i\in\mathbb{N}}$ is a sequence of measurable functions on a complete measure space X that converges pointwise almost everywhere to a function f. Show that f is measurable.

DEFINITION 6.3.13. Let X be a measure space. A function $f : X \longrightarrow [-\infty, \infty]$ that takes only finitely many values is called a *simple function*.

REMARK 6.3.14. We can write a simple function in the form $f(x) = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ for pairwise distinct α_i, where $A_i = f^{-1}(\alpha_i)$.

EXERCISE 6.3.15. Given that the α_i are pairwise distinct, show that a simple function $f(x) = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ is measurable iff each A_i is measurable.

Simple functions play an important role in every aspect of measure theory. The following exercise illustrates this in a direct way.

EXERCISE 6.3.16.

(i) If X is a measure space, show that any measurable function $f : X \longrightarrow [-\infty, \infty]$ is the pointwise limit everywhere of a sequence of measurable simple functions.

(ii) If f is a nonnegative measurable function, then the sequence in (i) can be taken to be nondecreasing.

(iii) Show that the sequence $(s_i)_{i\in\mathbb{N}}$ of simple functions converging to a measurable function $f : X \longrightarrow [-\infty, \infty]$ can be chosen such that s_i takes on values in the set $\{k \cdot 2^{-i} \mid k \in \mathbb{Z}\}$.

The following two theorems are among the most important theorems concerning measurable functions and their pointwise limits.

THEOREM 6.3.17 **(Egorov's theorem)**. *Suppose (X, μ) is a measure space and $\mu(X) < \infty$. If $(f_i : X \longrightarrow \mathbb{R})_{i\in\mathbb{N}}$ is a sequence of measurable functions converging pointwise almost everywhere to f, then, for any $\varepsilon > 0$, there is a subset $A \subseteq X$ with $\mu(X \setminus A) < \varepsilon$ such that $f_i \to f$ uniformly on A.*

Proof. Fix $j \in \mathbb{N}$. For $k \in \mathbb{N}$, let

$$C_{jk} = \bigcup_{i=k}^{\infty} \{x \in X \mid |f_i(x) - f(x)| \geq 2^{-j}\}.$$

Then $\bigcap_{k=1}^{\infty} C_{jk} \subseteq \{x \in X \mid \lim_{i\to\infty} f_i(x) \neq f(x)\}$, so $\mu(C_{jk}) \to 0$ as $k \to \infty$ (since the C_{jk} form a decreasing sequence). Choose $k(j)$ such that $\mu(C_{jk(j)}) < 2^{-j}\varepsilon$. Let $A = X \setminus \bigcup_{j=1}^{\infty} C_{jk(j)}$.

Exercise 6.3.18. Show that Egorov's theorem does not extend to sets of infinite measure.

Theorem 6.3.19 **(Lusin's theorem)**. *Let X be a metric space with a regular Borel measure μ such that $\mu(X) < \infty$. If $f : X \longrightarrow \mathbb{R}$ is a measurable function, then for any $\varepsilon > 0$, there is a closed subset $A \subseteq X$ with $\mu(X \setminus A) < \varepsilon$ such that $f|_A$ is continuous.*

Proof. Approximate f by a sequence of simple functions $(s_i)_{i\in\mathbb{N}}$ whose values are integer multiples of 2^{-i}, and choose disjoint closed sets $E_{ij} \subseteq s_i^{-1}(2^{-i}j)$ with $\mu(s_i^{-1}(2^{-i}j) \setminus E_{ij}) < 2^{-i-j}\varepsilon$. For each i, choose $k(i)$ such that $\mu\left(X \setminus \bigcup_{j=1}^{k(i)} E_{ij}\right) < 2^{-i}\varepsilon$. Let $D_i = \bigcup_{j=1}^{k(i)} E_{ij}$. Then, the restriction of s_i to D_i is continuous, and if we let $A = \bigcap_{i=1}^{\infty} D_i$, then $\mu(X \setminus A) < \varepsilon$, and the s_i converge uniformly on A to f. Hence, f is continuous on A.

Exercise 6.3.20. Show that the following converse of Lusin's theorem holds. Let X be a metric space with a regular Borel measure μ such that $\mu(X) < \infty$, and let f be a real-valued function on X. Suppose that for any $\varepsilon > 0$, there exists a closed set $A \subseteq X$ with $\mu(X \setminus A) < \varepsilon$ such that $f|_A$ is continuous. Show that f is measurable.

6.4. The Integral

Here we are, all dressed up and some place to go. It's time for integration! For a while, X will just be any old measure space, and μ will be any measure on X. We can define the integral relative to this measure without any specific reference to Lebesgue measure. However, our ultimate goal is a careful treatment of the Lebesgue integral, that is, integration with respect to Lebesgue measure.

Definition 6.4.1. If $s : X \longrightarrow [0, \infty]$ is a measurable simple function, $s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$, we define the *integral* of s on X with respect to μ to be

$$\int_X s \, d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i).$$

For a general measurable function $f : X \longrightarrow [0, \infty]$, we define

$$\int_X f \, d\mu = \sup_s \int_X s \, d\mu,$$

where the supremum is taken over all measurable simple functions $s : X \longrightarrow [0, \infty]$ such that $s \leq f$.

Note that, according to this definition, any nonnegative measurable function has an integral, which may well be $+\infty$. In this context, we employ the convention that $0 \cdot \infty = 0$ as we did in Definition 6.1.6, so that if $f \equiv 0$ and $\mu(X) = \infty$, then $\int_X f\, d\mu = 0$.

Some properties of the integral are very simple.

EXERCISE 6.4.2.

(i) Show that the integral is monotonic, that is, if f and g are measurable functions such that $0 \leq f \leq g$, then $0 \leq \int_X f\, d\mu \leq \int_X g\, d\mu$.

(ii) If $c \in [0, \infty)$, show that $\int_X cf\, d\mu = c \int_X f\, d\mu$.

We can integrate over smaller sets as well.

DEFINITION 6.4.3. If $f : X \longrightarrow [0, \infty]$ is a measurable function and E is a measurable subset of X, then $\int_E f\, d\mu$ is defined to be $\int_X \chi_E f\, d\mu$.

Note that, since E is measurable, $\chi_E f$ is also measurable (and nonnegative).

EXERCISE 6.4.4. Let $A, B \subseteq X$ be measurable sets such that $A \subseteq B$, and let $f : X \longrightarrow [0, \infty]$ be a measurable function. Show that $\int_A f\, d\mu \leq \int_B f\, d\mu$.

REMARK 6.4.5. We have been talking about the fact that the Riemann integral is inadequate for convergence purposes. We do know that if $(f_i)_{i \in \mathbb{N}}$ is a sequence of real-valued Riemann integrable functions on a closed, bounded rectangle $R \subseteq \mathbb{R}^n$ and if f_i converges uniformly to f, then f is Riemann integrable on R and $\int_R f_i \to \int_R f$ (see Exercise 5.10.55). This is about the best we can do with the Riemann integral. Other attempts are doomed to failure. First, we give an example and then an exercise.

EXAMPLE 6.4.6. Arrange the members of $\mathbb{Q} \cap [0, 1]$ in a sequence $(r_i)_{i \in \mathbb{N}}$, and let $f_i : [0, 1] \longrightarrow \mathbb{R}$ be given by

$$f_i(x) = \begin{cases} 1 & \text{if } x = r_j,\ j \leq i, \\ 0 & \text{elsewhere.} \end{cases}$$

Note that f_i is the characteristic function of $\{r_1, r_2, \ldots, r_i\}$. Then each f_i is Riemann integrable (being zero except at finitely many points). But the sequence $(f_i)_{i \in \mathbb{N}}$ is monotonically increasing and converges *pointwise* to the characteristic function of $\mathbb{Q} \cap [0, 1]$. This function is not continuous at *any* point and hence is not Riemann integrable. Nonetheless, $\lim_{i \to \infty} \int_0^1 f_i\, dx = 0$.

Now we proceed to the various convergence theorems concerning integrals. Keep in mind that we are still dealing only with nonnegative functions, but this deficiency will be corrected below. First, we consider the following ideas.

LEMMA 6.4.7. *Let $s : X \longrightarrow [0, \infty]$ be a measurable simple function on a measure space (X, μ). For any measurable set $C \subseteq X$, let $\nu(C) = \int_C s\, d\mu$. Then ν is a measure on X with the same σ-algebra as μ.*

Proof. Let $s = \sum_{i=1}^{k} \alpha_i \chi_{A_i}$. If $\{E_j\}_{j\in J}$ is a finite or countable collection of disjoint μ-measurable sets, we set $E = \bigcup_{j\in J} E_j$. Then, since all the terms involved are nonnegative,

$$\begin{aligned}\nu(E) &= \sum_{i=1}^{k} \alpha_i \mu(A_i \cap E) = \sum_{i=1}^{k} \alpha_i \sum_{j\in J} \mu(A_i \cap E_j) \\ &= \sum_{j\in J} \sum_{i=1}^{k} \alpha_i \mu(A_i \cap E_j) = \sum_{j\in J} \nu(E_j).\end{aligned}$$

Hence, ν is a measure.

Here is the first important convergence theorem.

THEOREM 6.4.8 **(Monotone Convergence Theorem)**. *Let* $(f_i : X \longrightarrow [0,\infty])_{i\in\mathbb{N}}$ *be an increasing sequence of measurable functions converging pointwise a.e. to* $f : X \longrightarrow [0,\infty]$. *Then* f *is measurable, and* $\int_X f_i \, d\mu \to \int_X f \, d\mu$ *as* $i \to \infty$.

Proof. Suppose first that f is finite except on a set of measure. By Proposition 6.3.9, we know that f is measurable. Let $m = \sup_{i\in\mathbb{N}} \int_X f_i \, d\mu$. Note that m may be ∞. Then $m \le \int_X f \, d\mu$, since $\int_X f_i \, d\mu \le \int_X f \, d\mu$ for every $i \in \mathbb{N}$. To prove the reverse inequality, let s be a measurable simple function with $0 \le s \le f$ and $s < \infty$. Let $c \in (0,1)$ be a constant, and define, for each $i \in \mathbb{N}$, $E_i = \{x \in X \mid f_i(x) \ge cs(x)\}$. Then each E_i is a measurable set, $E_1 \subseteq E_2 \subseteq \cdots$, and $\bigcup_{i=1}^{\infty} E_i = X$ (since $cs(x) < f(x)$ for almost every $x \in X$). Also

$$\int_X f_i \, d\mu \ge \int_{E_i} f_i \, d\mu \ge c \int_{E_i} s \, d\mu.$$

Letting $i \to \infty$ and applying Lemma 6.4.7 to s, we obtain $m \ge c \int_X s \, d\mu$ for every $c < 1$. Hence, $m \ge \int_X s \, d\mu$. But this is true for every measurable simple s with $0 \le s \le f$ and $s < \infty$, so $m \ge \int_X f \, d\mu$.

We leave the case where f is infinite on a set of positive measure to the reader.

The main goal of the next few theorems is to investigate the relations between limits and integrals. The Monotone Convergence Theorem shows that, under appropriate circumstances, limits and integrals can be interchanged. The next theorem shows that, with proper hypotheses, integrals and sums can be interchanged. Before proving this theorem, we present an example, an exercise, and a lemma.

EXAMPLE 6.4.9. Define $f_n : \mathbb{R} \longrightarrow [0,\infty]$ by

$$f_n(x) = \begin{cases} 4n^2 x & \text{for } 0 \le x \le 1/(2n), \\ -4n^2(x - 1/n) & \text{for } 1/(2n) \le x \le 1/n, \\ 0 & \text{otherwise.} \end{cases}$$

The pointwise limit of this sequence is the function $f = 0$. On the other hand, $\lim_{n\to\infty} \int_{\mathbb{R}} f_n \, dm = 1$, whereas $\int_{\mathbb{R}} f \, dm = 0$.

EXERCISE 6.4.10. Suppose s and t are nonnegative simple functions on X. Show that $\int_X (s+t)\, d\mu = \int_X s\, d\mu + \int_X t\, d\mu$.

LEMMA 6.4.11. *Let $f, g : X \longrightarrow [0,\infty]$ be measurable functions. Then $\int_X (f+g)\, d\mu = \int_X f\, d\mu + \int_X g\, d\mu$.*

Proof. Take sequences $(s_i)_{i\in\mathbb{N}}$, $(t_i)_{i\in\mathbb{N}}$ of measurable simple functions increasing pointwise to f, g, respectively. Then, $(s_i + t_i)_{i\in\mathbb{N}}$ increases pointwise to $f+g$. Now, the Monotone Convergence Theorem shows that $\int_X (f+g)\, d\mu = \int_X f\, d\mu + \int_X g\, d\mu$.

THEOREM 6.4.12. *If $(f_i : X \longrightarrow [0,\infty])_{i\in\mathbb{N}}$ is a sequence of measurable functions and $f = \sum_{i=1}^{\infty} f_i$, then $\int_X f\, d\mu = \sum_{i=1}^{\infty} \int_X f_i\, d\mu$.*

Proof. The case of finitely many measurable functions follows from Lemma 6.4.11.

If we set $g_i = f_1 + \cdots + f_i$, then $(g_i)_{i\in\mathbb{N}}$ increases pointwise to f, and the Monotone Convergence Theorem implies that $\int_X f\, d\mu = \lim_{i\to\infty} \int_X g_i\, d\mu = \sum_{i=1}^{\infty} \int_X f_i\, d\mu$.

We now generalize Lemma 6.4.7 to nonnegative measurable functions.

THEOREM 6.4.13. *Let (X,μ) be a measure space, and let $f : X \longrightarrow [0,\infty]$ be a measurable function. If C is a measurable set, we define $\nu(C) = \int_C f\, d\mu$. Then ν is a measure on X.*

Proof. Let $E_1, E_2, \ldots$ be a sequence of disjoint μ-measurable sets whose union is E. Then $\nu(E) = \int_X \chi_E f\, d\mu$, while $\nu(E_i) = \int_X \chi_{E_i} f\, d\mu$. But $\chi_E = \sum_{i\in I} \chi_{E_i}$, so just apply Theorem 6.4.12.

EXERCISE 6.4.14. What are the measurable sets with respect to the measure ν defined in Theorem 6.4.13?

EXERCISE 6.4.15. If $g : X \longrightarrow [0,\infty]$ is a measurable function, what is $\int_X g\, d\nu$, with ν as above?

THEOREM 6.4.16 **(Fatou's lemma)**. *If $(f_i : X \longrightarrow [0,\infty])_{i\in\mathbb{N}}$ is a sequence of measurable functions, then*

$$\int_X \left(\liminf_{i\to\infty} f_i \right) d\mu \leq \liminf_{i\to\infty} \int_X f_i\, d\mu.$$

Proof. Set $g_j = \inf_{i\geq j} f_i$. Then the sequence $(g_j)_{j\in\mathbb{N}}$ increases pointwise to $\liminf f_i$ as $j \to \infty$, so the Monotone Convergence Theorem yields $\lim_{j\to\infty} \int_X g_j\, d\mu = \int_X (\liminf_{i\to\infty} f_i)\, d\mu$. On the other hand, $\int_X g_j\, d\mu \leq \int_X f_j\, d\mu$ since $g_j \leq f_j$. Taking the $\liminf$ (which is a plain limit on the left, since the left side converges), we obtain the theorem.

This is what we always wanted to do—interchange sums and integrals, interchange limits and integrals, and all these wonderful processes of analysis that seemed forbidden in earlier days. Well, we can do them now, so let us go ahead and define integrals for functions that are not necessarily nonnegative.

DEFINITION 6.4.17. A measurable function $f : X \longrightarrow \mathbb{C} \cup \{\infty\}$ is said to be *integrable* if $\int_X |f|\, d\mu < \infty$. In this case, the integral of f is defined as follows. Let $f = u + iv$, where u and v are extended-real-valued (and measurable) functions. Let u^+, u^-, v^+, and v^- be as in Exercise 6.3.10. Then, we define

$$\int_X f\, d\mu = \int_X u^+\, d\mu - \int_X u^-\, d\mu + i \int_X v^+\, d\mu - i \int_X v^-\, d\mu.$$

EXERCISE 6.4.18. Let f be a bounded integrable function on a measure space (X, μ). Given $\varepsilon > 0$, show that there exist measurable simple functions s_1 and s_2 on X such that $s_1(x) \le f(x) \le s_2(x)$ for almost all $x \in X$ and $\int_X (s_2 - s_1)\, d\mu < \varepsilon$.

The following proposition shows that if f is integrable, then $\left|\int_X f\, d\mu\right| < \infty$.

PROPOSITION 6.4.19. *If $f : X \longrightarrow \mathbb{C} \cup \{\infty\}$ is integrable, then $\left|\int_X f\, d\mu\right| \le \int_X |f|\, d\mu$.*

Proof. First assume f is real-valued. Write $f = f^+ - f^-$, the difference of two nonnegative functions. Then

$$\left|\int_X (f^+ - f^-)\, d\mu\right| \le \int_X f^+\, d\mu + \int_X f^-\, d\mu = \int_X |f|\, d\mu.$$

The inequality for $f : X \longrightarrow \mathbb{C} \cup \{\infty\}$ follows from the exercise below.

EXERCISE 6.4.20. Extend the assertion that $\left|\int_X f\, d\mu\right| \le \int_X |f|\, d\mu$ to complex-valued functions. *Hint*: Choose a complex number c with $|c| = 1$ such that $\left|\int_X f\, d\mu\right| = \int_X cf\, d\mu$.

EXERCISE 6.4.21. Let R be a rectangle in $\mathbb{R}^n$. Suppose $f : R \longrightarrow \mathbb{R}$ is Riemann integrable on R. Show that the Riemann integral of f over R equals the Lebesgue integral of f over R. (Hence, the only way in which convergence theorems can fail for Riemann integration is when the limit function fails to be Riemann integrable.)

We now present what is probably the most important convergence theorem of them all, the Dominated Convergence Theorem.

THEOREM 6.4.22 **(Dominated Convergence Theorem)**. *Suppose $(f_i : X \longrightarrow \mathbb{C})_{i \in \mathbb{N}}$ is a sequence of measurable functions that converges pointwise to f and suppose $h : X \longrightarrow [0, \infty)$ is an integrable function such that $|f_i(x)| \le h(x)$ for all $x \in X$ and $i \in \mathbb{N}$. Then f is integrable on X, and $\int_X f_i\, d\mu \to \int_X f\, d\mu$ as $i \to \infty$.*

Proof. Set $g_i(x) = 2h(x) - |f_i(x) - f(x)|$. Then $(g_i)_{i\in\mathbb{N}}$ is a sequence of nonnegative measurable functions that converges pointwise to $2h$. We now apply Fatou's lemma:

$$\begin{aligned} 0 &\geq \int_X 2h\,d\mu - \liminf_{i\to\infty} \int_X (2h - |f_i - f|)\,d\mu \\ &= \limsup_{i\to\infty} \int_X |f_i - f|\,d\mu \\ &\geq \limsup_{i\to\infty} \left| \int_X f_i\,d\mu - \int_X f\,d\mu \right|. \end{aligned}$$

The Dominated Convergence Theorem tells us that if a sequence of functions is uniformly bounded and the domain is of finite measure, then everything is fine. The examples below show that if either of these two conditions is violated, limits and integrals cannot always be interchanged.

EXAMPLE 6.4.23. Consider the sequence of functions $(f_i : [0,1] \longrightarrow \mathbb{R})_{i\in\mathbb{N}}$ such that $f_i(0) = 2i$, f_i decreases linearly to $f_i(1/i) = 0$, and $f_i(x) = 0$ for $x \geq 1/i$. Then $\int_0^1 f_i\,dm = 1$, but $(f_i)_{i\in\mathbb{N}}$ converges pointwise to the function that is 0 except at 0, where it is ∞. The problem here is that any dominating function for the f_i looks essentially like $1/x$, which is not integrable. This is similar to Example 6.4.9.

EXAMPLE 6.4.24. Let $X = \mathbb{R}$ with Lebesgue measure. Define $f_n(x) : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$f_n(x) = \begin{cases} \frac{1}{1+|x|} & \text{for } |x| \leq n, \\ 0 & \text{otherwise.} \end{cases}$$

Then each f_n is integrable on $\mathbb{R}$, and the sequence $(f_n)_{n\in\mathbb{N}}$ converges uniformly to $f(x) = 1/(1 + |x|)$. Moreover, the f_n are uniformly bounded. Unfortunately f is not integrable.

We now present some elementary propositions about the integral that are interesting in their own right and will prove useful later.

EXERCISE 6.4.25. Suppose that X is a measure space and $E \subseteq X$ is a measurable set. If f is measurable on X, show that $\int_E |f(x)|\,d\mu = 0$ if and only if $f(x) = 0$ a.e. on E.

PROPOSITION 6.4.26 **(Absolute continuity of the integral)**. *Let (X, μ) be a measure space, and let f be an integrable function on X. Then, given $\varepsilon > 0$, there exists $\delta > 0$ such that if A is a measurable set and $\mu(A) < \delta$, then $\int_A |f|\,d\mu < \varepsilon$.*

Proof. Suppose there exists $\varepsilon > 0$ for which no corresponding δ exists. If $\delta_k = 2^{-k}$, there exists a set A_k such that $\mu(A_k) < \delta_k$ and $\int_{A_k} |f|\,d\mu > \varepsilon$.

Define a nested sequence of sets by $B_k = \bigcup_{i=k}^{\infty} A_i$ and let $B = \bigcap_{k=1}^{\infty} B_k$. Then

$$\mu(B_k) \leq \sum_{i=k}^{\infty} 2^{-i} = 2^{-k+1},$$

and hence $\mu(B) = 0$. Let $f_k = |f|\chi_{B_k}$ and note that $f_k \to |f|\chi_B$ pointwise as $k \to \infty$, the latter being zero almost everywhere. Also,

$$\int_X f_k \, d\mu \geq \int_{A_k} |f| \, d\mu > \varepsilon$$

for all k. But by the Dominated Convergence Theorem,

$$\int_X f_k \, d\mu \to \int_X |f|\chi_B \, d\mu = 0$$

by the previous exercise.

EXERCISE 6.4.27. Let X be a metric space, and let μ be a regular Borel measure on X. Then, given an integrable function $f : X \longrightarrow \mathbb{C}$ and $\varepsilon > 0$, show that there is a function $\phi \in C_c(X, \mathbb{C})$ such that $\int_X |f - \phi| \, d\mu < \varepsilon$. (Hint: See Lemma 2.7.27 .)

EXERCISE 6.4.28. With the notation of the previous exercise, show that if $X = \mathbb{R}^n$, we may replace $C_c(X, \mathbb{C})$ by $C_c^\infty(X, \mathbb{C})$ (see the project in Section 5.10.3).

PROPOSITION 6.4.29. *If $f : \mathbb{R}^n \longrightarrow \mathbb{C}$ is integrable, then $\int_{\mathbb{R}^n} |f(x+y) - f(x)| \, dm(x) \to 0$ as $y \to 0$.*

Proof. Fix $\varepsilon > 0$. Since f is integrable, we can take a continuous compactly supported function ϕ such that $\int_{\mathbb{R}^n} |f - \phi| \, dm < \varepsilon$. Let Q be a closed cube containing the support of ϕ. Let Q^* be a cube concentric with Q having diameter that is ten times the diameter of Q. Since ϕ is uniformly continuous, we can choose δ less than the diameter of Q such that $|\phi(x + y) - \phi(x)| < \varepsilon/m(Q^*)$ if $|y| < \delta$. If x is in the support of ϕ and $|y| < \delta$, then $x + y \in Q^*$, so

$$\int_{\mathbb{R}^n} |\phi(x+y) - \phi(x)| \, dm(x) = \int_{Q^*} |\phi(x+y) - \phi(x)| \, dm(x) < \varepsilon.$$

Hence,

$$\begin{aligned}
&\int_{\mathbb{R}^n} |f(x+y) - f(x)| \, dm(x) \\
&\quad \leq \int_{\mathbb{R}^n} |f(x+y) - \phi(x+y)| \, dm(x) \\
&\qquad + \int_{\mathbb{R}^n} |f(x) - \phi(x)| \, dm(x) + \int_{\mathbb{R}^n} |\phi(x+y) - \phi(x)| \, dm(x) \\
&\quad < 3\varepsilon
\end{aligned}$$

if $|y| < \delta$.

There is an additional convergence theorem that is of considerable interest in integration theory.

THEOREM 6.4.30 **(Vitali Convergence Theorem)**. *Let (X, μ) be a measure space such that $\mu(X) < \infty$. Suppose we have a sequence $(f_n)_{n\in\mathbb{N}}$ of integrable functions such that $f_n \to f$ a.e. as $n \to \infty$ and $|f(x)| < \infty$ a.e. The following statements are equivalent:*

(i) *f is integrable, and for all measurable sets A, $\int_A f_n \, d\mu \to \int_A f \, d\mu$ as $n \to \infty$;*

(ii) *the integrals $\int |f_n| \, d\mu$ are uniformly absolutely continuous, that is, for any $\varepsilon > 0$ there exists $\delta > 0$ such that if A is a measurable set and $\mu(A) < \delta$, then for any $n \in \mathbb{N}$, we have $\int_A |f_n| \, d\mu < \varepsilon$.*

Proof. First we show that (ii) implies (i). Suppose that (ii) holds and choose $\rho < 1$ such that, for all measurable A, if $\mu(A) < \rho$, then $\int_A |f_n| \, d\mu < 1$. By Egorov's theorem (Theorem 6.3.17), since $f_n \to f$ a.e., there exists a measurable set B such that $\mu(B) < \rho$ and $f_n \to f$ uniformly on $X \setminus B$. Choose n_0 so that $|f - f_{n_0}| < 1$ on $X \setminus B$. Then $|f| < |f_{n_0}| + 1$ on $X \setminus B$ and

$$\int_{X\setminus B} |f| \, d\mu \leq \int_{X\setminus B} (|f_{n_0}| + 1) \, d\mu.$$

Now, using Fatou's lemma on B, we have

$$\int_B |f| \, d\mu = \int_B \liminf_n |f_n| \, d\mu \leq \liminf_n \int_B |f_n| \, d\mu \leq 1.$$

Hence, f is integrable on X.

We claim that $\int_X |f - f_n| \, d\mu \to 0$ as $n \to \infty$. Suppose $\varepsilon > 0$. By Proposition 6.4.26, there is a $\delta_1 > 0$ such that if A is a measurable set and $\mu(A) < \delta_1$, then $\int_A |f| \, d\mu < \varepsilon$. By assumption, there exists $\delta_2 > 0$ such that if A is a measurable set with $\mu(A) < \delta_2$, then, for all $n \in \mathbb{N}$, $\int_A |f_n| \, d\mu < \varepsilon$. We apply Egorov's theorem again to find a measurable set B' such that $\mu(B') < \min(\delta_1, \delta_2)$ and $f_n \to f$ uniformly on $X \setminus B'$. Choose n_1 such that $|f - f_n| < \varepsilon$ on $X \setminus B'$ for $n \geq n_1$. Then for $n \geq n_1$, we have

$$\int_X |f - f_n| \, d\mu \leq \int_{B'} |f| \, d\mu + \int_{B'} |f_n| \, d\mu + \int_{X\setminus B'} |f - f_n| \, d\mu < (2 + \mu(X))\varepsilon.$$

This proves (i).

Now we show that (i) implies (ii). Assume (i) holds. Without loss of generality assume that $f_n \to 0$ a.e. and $\int_A f_n \, d\mu \to 0$ for all measurable sets A. Furthermore, it is sufficient to show that $\int_X f_n \, d\mu \to 0$. This is because, for a fixed n, the integrals $\int |f_n| \, d\mu$ are absolutely continuous. Hence, for a given $\varepsilon > 0$ there exists $\delta_k > 0$ so that if $\mu(A) < \delta_k$, then $\int_A f_k \, d\mu < \varepsilon$. Thus, given $\varepsilon > 0$, to show that we can find a uniform $\delta > 0$, we choose an n_1 such that $\int_X f_n \, d\mu < \varepsilon$ for $n > n_1$. Now take $\delta = \min(\delta_1, \ldots, \delta_{n_1})$.

We now prove $\int_X f_n \, d\mu \to 0$. Assume not. Then, for some $\delta > 0$ there is an infinite set $S \subseteq \mathbb{N}$ such that $\int_X |f_n| \, d\mu > \delta$ for all $n \in S$. Now for $k \in \mathbb{N}$

let

$$E_k = \{x \in X \mid |f_n(x)| \leq k \text{ for all } n\}.$$

Clearly $E_k \subseteq E_{k+1}$ for all $k \in \mathbb{N}$ and $\mu\left(X \setminus \bigcup_k E_k\right) = 0$. Hence $\mu(X \setminus E_k) \to 0$ as $k \to \infty$ by Proposition 6.1.11.

Let $k_1 = 1$. By definition, $|f_n| < 1$ on E_{k_1}, and $f_n \to 0$ a.e. By the Dominated Convergence Theorem, $\int_{E_{k_1}} f_n \, d\mu \to 0$ as $n \to \infty$. Now choose $n_1 \in S$ such that $\int_{E_{k_1}} |f_{n_1}| \, d\mu < \varepsilon$. Then

$$\int_{X \setminus E_{k_1}} |f_{n_1}| \, d\mu \geq \int_X |f_{n_1}| \, d\mu - \int_{E_{k_1}} |f_{n_1}| \, d\mu > \delta - \varepsilon.$$

Now choose $k_2 > k_1$ such that $\int_{X \setminus E_{k_2}} f_{n_1} \, d\mu < \varepsilon$. We can do this by absolute continuity of the integral and the fact that $\mu(X \setminus E_k) \to 0$ as $k \to \infty$. Again by the Dominated Convergence Theorem we can choose $n_2 \in S$ with $n_2 > n_1$ and $\int_{E_{k_2}} |f_{n_2}| \, d\mu < \varepsilon$. Again as before $\int_{X \setminus E_{k_2}} |f_{n_2}| \, d\mu > \delta - \varepsilon$. We may induct to find sequences $(k_r)_{r \in \mathbb{N}}$ in $\mathbb{N}$ and $(n_r)_{r \in \mathbb{N}}$ in S such that for all r,

$$\int_{X \setminus E_{k_{r+1}}} |f_{n_r}| \, d\mu < \varepsilon, \qquad \int_{E_{k_r}} |f_{n_r}| \, d\mu < \varepsilon, \qquad \int_{X \setminus E_{k_r}} |f_{n_r}| \, d\mu > \delta - \varepsilon.$$

Then for $r \in \mathbb{N}$, we have

$$\int_{E_{k_{r+1}} \setminus E_{k_r}} |f_{n_r}| \, d\mu \geq \int_{X \setminus E_{k_r}} |f_{n_r}| \, d\mu - \int_{X \setminus E_{k_{r+1}}} |f_{n_r}| \, d\mu > \delta - 2\varepsilon.$$

Now let $E = \coprod_{r \in \mathbb{N}} E_{k_{r+1}} \setminus E_{k_r}$. Then for $r \in \mathbb{N}$,

$$\int_E |f_{n_r}| \, d\mu \geq \int_{E_{k_{r+1}} \setminus E_{k_r}} |f_{n_r}| \, d\mu - \int_{E_{k_{r-1}}} |f_{n_r}| \, d\mu - \int_{X \setminus E_{k_{r+1}}} |f_{n_r}| \, d\mu.$$

We find that

$$\int_E |f_{n_r}| \, d\mu \geq \delta - 4\varepsilon > \delta/2,$$

for ε sufficiently small. This contradicts the fact that $f_n \to 0$ a.e., so we see that (i) implies (ii).

6.5. L^p Spaces

Among the most important spaces for analysis are the spaces $L^p(\mu)$ of p-th power integrable functions. For us, most applications will be in $X = \mathbb{R}^n$ where μ is Lebesgue measure. Of course, we have already encountered these L^p spaces in Chapter 4 in the case where $X = \mathbb{N}$ and μ is the counting measure. In that case $L^p(\mu)$ was denoted $\ell^p(\mathbb{C})$.

DEFINITION 6.5.1. Let (X, μ) be a measure space. For an exponent p, $1 \leq p < \infty$, let $L^p(\mu)$ be the set of measurable functions $f : X \longrightarrow \mathbb{C}$ such that

$$\|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{1/p} < \infty.$$

The reader might object by saying that, even though we have used the "norm" symbol $\| \cdot \|_p$, this is not really a norm, since if $f = 0$ almost everywhere, then $\|f\|_p = 0$. We prefer to ignore this fact. One should really define equivalence classes of functions that are equal almost everywhere and then define the norm on these equivalence classes. However, for the purposes of analysis, it makes no difference whatsoever, and we will use the symbol $L^p(\mu)$ for both the functions themselves and the equivalence classes.

There is also the case $p = \infty$.

DEFINITION 6.5.2. Let (X, μ) be a measure space. We say that a measurable function $f : X \longrightarrow \mathbb{C}$ is *essentially bounded* if there exists a positive constant C such that $\mu(\{x \in X \mid |f(x)| > C\}) = 0$. If f is essentially bounded, we define $\|f\|_\infty$ (the *essential supremum* of f) to be the infimum over all such constants C. We denote by $L^\infty(\mu)$ the space of all essentially bounded functions.

Again, as in the case of $L^p(\mu)$, $L^\infty(\mu)$ is really a space of equivalence classes. This will cause no difficulty as long as the reader does not confuse supremum and essential supremum.

Most of this section will be devoted to proving that the $L^p(\mu)$ spaces are Banach spaces ($L^2(\mu)$ is actually a Hilbert space). We require some classical inequalities to do this. As remarked above, it is important to remember when dealing with these spaces that writing $f \in L^p(\mu)$ really determines the function f only up to a set of measure zero.

EXERCISE 6.5.3. If $1 \leq p < q \leq \infty$, find a function f that is in $L^p(\mathbb{R}, m)$ but not in $L^q(\mathbb{R}, m)$. Also, find a function that is in $L^q(\mathbb{R}, m)$ but not in $L^p(\mathbb{R}, m)$.

DEFINITION 6.5.4. If p is a real number, $1 < p < \infty$, we denote the exponent *conjugate* to p by $1/q = 1 - 1/p$, so that $\frac{1}{p} + \frac{1}{q} = 1$. If $p = 1$, we set $q = \infty$, and vice versa. Note that $p = 2$ if and only if $q = 2$. Ordinarily, when the letters p and q are used together, they represent conjugate exponents.

In proving the triangle inequality for the spaces $L^p(\mu)$, we proceed as we did in Chapter 2 and first prove Hölder's inequality.

PROPOSITION 6.5.5 **(Hölder's inequality)**. *Let f, $g : X \longrightarrow [0, \infty]$ be measurable functions, and take $1 < p < \infty$. Then*

$$\int_X fg \, d\mu \leq \left(\int_X f^p \, d\mu \right)^{1/p} \left(\int_X g^q \, d\mu \right)^{1/q}.$$

Note that this includes the Cauchy-Schwarz inequality as the case $p = 2$.

Proof. Set $A = \left(\int_X f^p\,d\mu\right)^{1/p}$, $B = \left(\int_X g^q\,d\mu\right)^{1/q}$. If either A or B is 0 or ∞, then the inequality is obvious. Otherwise, divide f and g by A and B, respectively, to reduce the right side of the inequality to 1. Now let $X' = \{x \in X \mid f(x)g(x) > 0\}$. We have $0 < f(x) < \infty$ and $0 < g(x) < \infty$ almost everywhere on X'. For $x \in X'$, let $s = p\log f(x)$, $t = q\log g(x)$. Since the function exp is convex, we have, for almost every $x \in X'$,

$$f(x)g(x) = e^{s/p+t/q} \le e^s/p + e^t/q = f(x)^p/p + g(x)^q/q,$$

since $1/p + 1/q = 1$. Integrating this inequality, we get

$$\int_X fg\,d\mu \le 1.$$

COROLLARY 6.5.6. *Let f, $g : X \longrightarrow \mathbb{C}$ be measurable functions, and let $1 < p < \infty$. Then*

$$\int_X |fg|\,d\mu \ \le \left(\int_X |f|^p\,d\mu\right)^{1/p}\left(\int_X |g|^q\,d\mu\right)^{1/q}.$$

EXERCISE 6.5.7. Show that Hölder's inequality can be generalized as follows. Let $1 < p_k < \infty$ for $k = 1, \ldots, N$ satisfy $\sum_{k=1}^N 1/p_k = 1$ and let $f_k : X \longrightarrow \mathbb{C}$ be measurable functions. Then

$$\int_X |f_1 f_2 \cdots f_N| d\mu \le \|f_1\|_{p_1}\|f_2\|_{p_2}\cdots\|f_N\|_{p_N}.$$

Hölder's inequality is an interesting inequality in its own right and is also extremely useful (if not essential) to prove the next result.

PROPOSITION 6.5.8. *Let f, $g : X \longrightarrow [0,\infty]$ be measurable functions, and let $1 < p < \infty$. Then*

$$\left(\int_X (f+g)^p\,d\mu\right)^{1/p} \le \left(\int_X f^p\,d\mu\right)^{1/p} + \left(\int_X g^p\,d\mu\right)^{1/p}.$$

Proof. If $\int_X (f+g)^p\,d\mu = 0$, the result is obvious. If $\int_X (f+g)^p\,d\mu = \infty$, then we observe that $(f+g)^p \le 2^p(f^p + g^p)$, so either $\int_X f^p\,d\mu = \infty$ or $\int_X g^p\,d\mu = \infty$.

Otherwise, write

$$(f+g)^p = f(f+g)^{p-1} + g(f+g)^{p-1},$$

and apply Hölder's inequality with $q = p/(p-1)$ to each of the two terms. Adding the resulting inequalities, we get

$$\int_X (f+g)^p\,d\mu \le \left(\int_X (f+g)^p\,d\mu\right)^{1/q}\left(\left(\int_X f^p\,d\mu\right)^{1/p} + \left(\int_X g^p\,d\mu\right)^{1/p}\right).$$

Dividing this inequality by $\left(\int_X (f+g)^p\,d\mu\right)^{1/q}$, we get the desired inequality.

COROLLARY 6.5.9 **(Minkowski's inequality)**. *Let $f, g : X \longrightarrow \mathbb{C}$ be measurable functions, and let $1 < p < \infty$. Then*

$$\left(\int_X |f+g|^p \, d\mu\right)^{1/p} \leq \left(\int_X |f|^p \, d\mu\right)^{1/p} + \left(\int_X |g|^p \, d\mu\right)^{1/p}.$$

This is, of course, the triangle inequality for the L^p norm. Note that the triangle inequality for L^1 is obvious.

For $1 < p < \infty$, the above discussion yields the following proposition with the usual caveat about sets of measure zero.

PROPOSITION 6.5.10. *The space $L^p(\mu)$ is a normed linear space for all p, $1 \leq p < \infty$.*

EXERCISE 6.5.11. Prove that $L^\infty(\mu)$ is a normed linear space. As usual, watch out for sets of measure zero.

EXERCISE 6.5.12. Find conditions on the measure μ that ensure that, when $r < s$,

(i) $L^r(\mu) \subseteq L^s(\mu)$;
(ii) $L^s(\mu) \subseteq L^r(\mu)$;
(iii) neither inclusion holds.

EXERCISE 6.5.13. Determine whether the following statement is true or false: if (X, μ) is a measure space with $\mu(X) < \infty$, then we have $L^\infty(\mu) = \bigcap_{1 \leq p < \infty} L^p(\mu)$.

THEOREM 6.5.14. *For all p, $1 \leq p \leq \infty$, $L^p(\mu)$ is a Banach space.*

Proof. First suppose that $1 \leq p < \infty$. The completeness proof for $p = \infty$ will be simpler. Let $(f_i)_{i \in \mathbb{N}}$ be a Cauchy sequence in $L^p(\mu)$. Without loss of generality, we may assume that $\|f_{i+1} - f_i\|_p < 2^{-i}$ (see Exercise 1.6.13(ii)). Set $g_j = \sum_{i=1}^{j} |f_{i+1} - f_i|$. We observe that g_j is finite almost everywhere because f_i is finite almost everywhere. We set $g = \lim_{j \to \infty} g_j$ almost everywhere. Then g is measurable. By Minkowski's inequality, $\|g_j\|_p < 1$ for every j. Thus, $\|g\|_p \leq 1$ by Fatou's lemma. This means that $g(x) < \infty$ almost everywhere, and hence the limit of $(f_i(x))_{i \in \mathbb{N}}$ exists for almost every x. Set $f(x) = \lim_{i \to \infty} f_i(x)$, when this limit exists, and $f(x) = 0$ on the remaining set of measure zero. We shall verify that $f_i \to f$ in the L^p norm.

Choose $\varepsilon > 0$ and $N \in \mathbb{N}$ such that $\|f_i - f_j\|_p < \varepsilon$ if $i, j > N$. By Fatou's lemma,

$$\int_X |f - f_j|^p \, d\mu \leq \liminf_{i \to \infty} \int_X |f_i - f_j|^p \, d\mu \leq \varepsilon^p.$$

Hence, $f - f_j \in L^p(\mu)$, $f \in L^p(\mu)$, and $\|f - f_j\|_p \to 0$ as $j \to \infty$.

In the case $p = \infty$, let $(f_i)_{i \in \mathbb{N}}$ be a Cauchy sequence. Also, let $A_i = \{x \mid |f_i(x)| > \|f_i\|_\infty\}$, and let $B_{ij} = \{x \mid |f_i(x) - f_j(x)| > \|f_i - f_j\|_\infty\}$. Then every A_i and B_{ij} has measure zero, so $E = (\bigcup_i A_i) \cup \left(\bigcup_{i,j} B_{ij}\right)$ has measure zero. On ${}^cE$, the f_i are bounded and converge uniformly to a bounded

function f. Setting $f(x) = 0$ for $x \in E$, we get an L^∞ function, which is the L^∞ limit of the f_i.

EXERCISE 6.5.15. Let X be a metric space, and let μ be a regular Borel measure on X. When $1 \leq p < \infty$, show that $C_c(X, \mathbb{C})$ (the space of continuous complex-valued functions of compact support) is dense in $L^p(\mu)$. (Of course, the elements of $L^p(\mu)$ are equivalence classes, and the elements of $C_c(X, \mathbb{C})$ are functions. But we can identify $C_c(X, \mathbb{C})$ with a subspace of $L^p(\mu)$, since each class in $L^p(\mu)$ contains at most one element of $C_c(X, \mathbb{C})$.)

EXERCISE 6.5.16. If $X = \mathbb{R}^n$, we can do better than just continuous functions. Show that the subspace $C_c^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n, m)$, where m is Lebesgue measure and $p < \infty$.

EXERCISE 6.5.17. Show that the completion of $C_c(X, \mathbb{C})$ in the L^∞ norm (again, identifying $C_c(X, \mathbb{C})$ with a subspace of $L^\infty(\mu)$) is $C_0(X, \mathbb{C})$, the set of continuous functions f on X that vanish at infinity in the sense that, for any $\varepsilon > 0$, there is a compact set $K \subseteq X$ such that $|f(x)| < \varepsilon$ for $x \notin K$.

6.6. Fubini's Theorem

All integrals in this sections are taken with respect to Lebesgue measure. We will usually write $\int_{\mathbb{R}^n} f(x)\, dx$ in place of $\int_{\mathbb{R}^n} f\, dm$ for an integral on Euclidean space. This will be especially handy when we are dealing with variables coming from several Euclidean spaces. Let $X = \mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^h$. We denote a point of $\mathbb{R}^n$ by $z = (x, y)$, with $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^h$. For a function f defined on $\mathbb{R}^n$ and a fixed $y \in \mathbb{R}^h$, let f_y be the section of f defined on $\mathbb{R}^k$ at y, that is, $f_y(x) = f(x, y)$. For $A \subseteq \mathbb{R}^n$, let A_y be the section of A at y, that is, $A_y = \{x \mid (x, y) \in A\}$. If f_y is integrable, set $F(y) = \int_{\mathbb{R}^k} f_y(x)\, dx$.

We first prove Fubini's theorem for nonnegative functions.

THEOREM 6.6.1 **(Fubini)**. *If $f : \mathbb{R}^n \longrightarrow [0, \infty]$ is measurable, then f_y is measurable for almost every y, so F is almost everywhere defined. The function F is measurable on $\mathbb{R}^h$, and*

$$\int_{\mathbb{R}^h} F(y)\, dy = \int_{\mathbb{R}^n} f(z)\, dz.$$

This theorem is proved in a sequence of exercises. Most of the exercises are straightforward and follow directly from material developed earlier.

EXERCISE 6.6.2. Let $R \subseteq \mathbb{R}^n$ be a rectangle (with sides parallel to the coordinate axes). Show that Fubini's theorem holds for $f = \chi_R$, the characteristic function of R.

EXERCISE 6.6.3. Let $V \subseteq \mathbb{R}^n$ be a bounded open set. Show that Fubini's theorem holds for $f = \chi_V$. This follows easily from Lemma 6.2.41 and the previous exercise.

EXERCISE 6.6.4. Let $K \subseteq \mathbb{R}^n$ be a compact set. Show that Fubini's theorem holds for $f = \chi_K$.

EXERCISE 6.6.5. Let $E = \bigcup_{i=1}^{\infty} K_i$ be the countable union of an increasing sequence $K_1 \subseteq K_2 \subseteq \cdots$ of compact sets. Using the Monotone Convergence Theorem, show that Fubini's theorem holds for $f = \chi_E$.

EXERCISE 6.6.6. Let $E = \bigcap_{i=1}^{\infty} V_i$ be the countable intersection of a decreasing sequence $V_1 \supseteq V_2 \supseteq \cdots$ of open sets. Using the previous exercise, show that Fubini's theorem holds for $f = \chi_E$.

EXERCISE 6.6.7. Let $E \subseteq \mathbb{R}^n$ be a bounded measurable set. Show that Fubini's theorem holds for $f = \chi_E$. *Hint*: Use the regularity of Lebesgue measure and Exercise 6.4.25.

EXERCISE 6.6.8. Let $E \subseteq \mathbb{R}^n$ be any measurable set. Using the Monotone Convergence Theorem, show that Fubini's theorem holds for $f = \chi_E$.

EXERCISE 6.6.9. Let f be a nonnegative measurable simple function. Show that Fubini's theorem holds for f.

Congratulations! Now you can finish the proof of Fubini's theorem for nonnegative measurable functions.

THEOREM 6.6.10 **(Fubini)**. *Let $f \in L^1(\mathbb{R}^n)$. Then $f_y \in L^1(\mathbb{R}^k)$ for almost every $y \in \mathbb{R}^h$. Also, the function*

$$F(y) = \int_{\mathbb{R}^k} f_y(x)\,dx$$

is an element of $L^1(\mathbb{R}^h)$, and

$$\int_{\mathbb{R}^h} F(y)\,dy = \int_{\mathbb{R}^n} f(z)\,dz.$$

EXERCISE 6.6.11. Prove it. (This is an easy exercise in the use of positive and negative parts.)

Here is a nice application of Fubini's theorem.

THEOREM 6.6.12 **(Minkowski's integral inequality)**. *If $1 \leq p < \infty$ and $f \in L^1(\mathbb{R}^n \times \mathbb{R}^n) \cap L^p(\mathbb{R}^n \times \mathbb{R}^n)$, then*

$$\left(\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} f(x,y)\,dy\right|^p dx\right)^{1/p} \leq \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |f(x,y)|^p\,dx\right)^{1/p} dy.$$

Proof. The proof for $p = 1$ follows from Fubini's theorem and Proposition 6.4.19.

Now take $p > 1$. Suppose first that f is real-valued and nonnegative. For $x \in \mathbb{R}^n$ define

$$F(x) = \int_{\mathbb{R}^n} f(x,y)\,dy.$$

If $F(x) = 0$ almost everywhere, then Minkowski's integral inequality is trivial, so we will suppose that F is not almost everywhere 0. By Fubini's

theorem, F is measurable. If $F \in L^p(\mathbb{R}^n)$, we can apply Fubini's theorem as follows:

$$\begin{aligned}\int_{\mathbb{R}^n} F(x)^p\,dx &= \int_{\mathbb{R}^n} \left(F(x)^{p-1} \int_{\mathbb{R}^n} f(x,y)\,dy \right) dx \\ &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} F(x)^{p-1} f(x,y)\,dx \right) dy.\end{aligned}$$

By Hölder's inequality, we have

$$\begin{aligned}&\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} F(x)^{p-1} f(x,y)\,dx \right) dy \\ &\quad\le \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x,y)^p\,dx \right)^{1/p} \left(\int_{\mathbb{R}^n} F(x)^p\,dx \right)^{\frac{p-1}{p}} dy \\ &\quad= \left(\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x,y)^p\,dx \right)^{1/p} dy \right) \left(\int_{\mathbb{R}^n} F(x)^p\,dx \right)^{\frac{p-1}{p}}.\end{aligned}$$

Now divide by

$$\left(\int_{\mathbb{R}^n} F(x)^p\,dx \right)^{\frac{p-1}{p}}$$

to obtain the desired inequality.

In case F is not necessarily in $L^p(\mathbb{R}^n)$, let $f_j = \chi_{[-j,j]^n \times [-j,j]^n} \min\{f, j\}$ be the truncation of f at j, let

$$F_j(x) = \int_{\mathbb{R}^n} f_j(x,y)\,dy,$$

and apply the Monotone Convergence Theorem.

Finally, for general complex-valued f, the result follows from the inequality

$$\left| \int_{\mathbb{R}^n} f(x,y)dy \right| \le \int_{\mathbb{R}^n} |f(x,y)|dy.$$

We now give a simple application of Minkowski's integral inequality that is useful in the study of integral operators.

Definition 6.6.13. Suppose that $f \in L^1(\mathbb{R}^n)$. Define an operator T_f on $L^p(\mathbb{R}^n)$, $1 \le p < \infty$, by $T_f(g) = f * g$ for $g \in L^p(\mathbb{R}^n)$, where

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y) g(x-y) dy.$$

Corollary 6.6.14 **(To Minkowski's integral inequality)**. *For $f \in L^1(\mathbb{R}^n)$, T_f is a bounded operator on $L^p(\mathbb{R}^n)$, and $\|T_f\| \le \|f\|_1$.*

Proof. If $p = 1$, this is a simple consequence of Fubini's theorem. If $p > 1$, then by Minkowski's integral inequality,

$$\begin{aligned}\left(\int_{\mathbb{R}^n} |(f*g)(x)|^p dx\right)^{1/p} &\leq \left(\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} f(y)g(x-y)dy\right|^p dx\right)^{1/p} \\ &\leq \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |f(y)|^p |g(x-y)|^p dx\right)^{1/p} dy \\ &= \int_{\mathbb{R}^n} |f(y)| \left(\int_{\mathbb{R}^n} |g(x-y)|^p dx\right)^{1/p} dy.\end{aligned}$$

By translation invariance, the inner integral is equal to

$$\int_{\mathbb{R}^n} |g(x)|^p dx,$$

so we get

$$\begin{aligned}\|T_f(g)\|_p &= \left(\int_{\mathbb{R}^n} |(f*g)(x)|^p dx\right)^{1/p} \\ &\leq \left(\int_{\mathbb{R}^n} |f(y)|dy\right)\left(\int_{\mathbb{R}^n} |g(x)|^p dx\right)^{1/p} = \|f\|_1 \|g\|_p.\end{aligned}$$

EXERCISE 6.6.15. Show that if $p = 1$, then the norm of the operator T_f is equal to $\|f\|_1$.

EXERCISE 6.6.16. Now let $f \in L^q(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$ and $g \in L^p(\mathbb{R}^n)$, and choose r such that

$$\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1.$$

Show that $f * g \in L^r(\mathbb{R}^n)$ and $\|f*g\|_r \leq \|f\|_q \|g\|_p$. Exercise 6.5.7 may be useful.

From the above proof, we know that T_f is a bounded linear operator on $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$. The norm of this operator does not exceed $\|f\|_1$. The actual determination of the norm is stated as a challenge problem at the beginning of the next chapter.

There is a general form of Fubini's theorem for the product of two σ-finite measure spaces.

DEFINITION 6.6.17. Let $\mathfrak{M}$ and $\mathfrak{N}$ be σ-algebras. Their product is defined to be the σ-algebra $\mathfrak{M} \times \mathfrak{N}$ generated by sets of the form $A \times B$ where $A \in \mathfrak{M}$ and $B \in \mathfrak{N}$.

Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces. For a set $E \in \mathfrak{M} \times \mathfrak{N}$ and points $x \in X$ and $y \in Y$, we define

$$E_x = \{y \in Y \mid (x, y) \in E\}$$

and

$$E^y = \{x \in X \mid (x, y) \in E\}.$$

THEOREM 6.6.18. *The sets E_x and E^y are measurable, and we have*

$$\int_X \nu(E_x)\, d\mu = \int_Y \mu(E^y)\, d\nu.$$

The assumption of σ-finiteness is important in the proof of this theorem.

DEFINITION 6.6.19. The product measure is defined to be the measure on $(X \times Y, \mathfrak{M} \times \mathfrak{N})$ defined by

$$(\mu \times \nu)(E) = \int_X \nu(E_x) d\mu = \int_Y \mu(E^y) d\nu.$$

THEOREM 6.6.20 **(Fubini's theorem)**. *Let $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite measure spaces, and let $f \in L^1(X \times Y)$. Then*

(i) *the function f_x defined by $f_x(y) = f(x, y)$ lies in $L^1(Y)$ for almost all $x \in X$,*
(ii) *the function f^y defined by $f^y(x) = f(x, y)$ lies in $L^1(X)$ for almost all $y \in Y$,*
(iii) *the function $x \mapsto \int_Y f_x d\nu$ lies in $L^1(X)$ for almost all $y \in Y$,*
(iv) *the function $y \mapsto \int_X f^y d\mu$ lies in $L^1(Y)$ for almost all $x \in X$, and*
(v) *we have*

$$\int_{X \times Y} f d(\mu \times \nu) = \int_X \left(\int_Y f_x d\nu \right) d\mu = \int_Y \left(\int_X f^y d\mu \right) d\nu.$$

The proofs of the above theorems can be found in [**25**], [**23**], [**7**], and [**4**].

6.7. Change of Variables in Integration

In this section, we derive a formula for "change of variables" in the Lebesgue integral. The difficulties involved in proving this theorem actually appear in any theory of integration and are not necessarily related to Lebesgue integration itself. The theorem can be stated precisely as follows.

THEOREM 6.7.1 **(Change of variables in $\mathbb{R}^n$)**. *Let $U \subseteq \mathbb{R}^n$ be open and let $\phi : U \longrightarrow \phi(U) \subseteq \mathbb{R}^n$ be a bijective function such that ϕ and ϕ^{-1} are C^1 functions. Let A be a measurable subset of U and let $B = \phi(A)$. Then a measurable function f on B is integrable iff $f \circ \phi$ is integrable on A, and in this case we have*

$$\int_B f(y)\, dy = \int_A f(\phi(x)) |J(x)|\, dx,$$

where $J(x) = \det D\phi(x)$ is the determinant of the Jacobian matrix of ϕ (see Definition 5.3.5).

We first prove this theorem for a linear transformation $\phi \in GL_n(\mathbb{R})$.

LEMMA 6.7.2. *Let $\phi \in GL_n(\mathbb{R})$, let A be a measurable subset of $\mathbb{R}^n$, and let $B = \phi(A)$. Then, a measurable function f on B is integrable iff $f \circ \phi$ is integrable on A, and in this case we have*

$$\int_B f(y)\,dy = \int_A f(\phi(x))|J(x)|dx.$$

Proof. We need only check that for any measurable set $A \subseteq \mathbb{R}^n$ with $m(A) < \infty$, we have $m(\phi(A)) = |\det(\phi)|m(A)$. Given this, the result follows from the density of simple functions.

Since every element ϕ of $GL_n(\mathbb{R})$ can be written as a product $\phi = kau$, where $k \in O_n(\mathbb{R})$, a is a diagonal matrix with positive entries, and u is an upper-triangular unipotent matrix

$$n = \begin{pmatrix} 1 & * & \cdots & * \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & 1 \end{pmatrix},$$

it follows from Theorems 6.2.38 and 6.2.40 and the fact that $\det\phi = \det a$ that we need only consider the action of an upper-triangular unipotent matrix on a measurable set.

We prove this for $n = 3$ using Fubini's theorem. The result for general n will then be immediately obvious.

$$\iiint_{\mathbb{R}^3} f\left[\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right] dx\,dy\,dz$$
$$= \iiint_{\mathbb{R}^3} f(x + ay + bz, y + cz, z)\,dx\,dy\,dz.$$

Since dx, dy, and dz are invariant under additive translation, we can successively replace $x + ay + bz$ by x and $y + cz$ by y, and z is not affected.

We now proceed to the general proof of change of variables. We first prove a lemma about the change of measure under the action of a C^1 function with a nonvanishing Jacobian. This process takes several steps.

LEMMA 6.7.3. *With notation as in Theorem* 6.7.1*, if A is a rectangle, then $B = \phi(A)$ is measurable, and $m(B) = \int_A |J(x)|\,dx$.*

Proof. Subdivide A into subrectangles $A_1, \ldots, A_N$ that are similar to A and congruent to each other, and let c_i denote the center of A_i. Let T^{c_i} be the first-order approximation to f, as in Section 5.9. Further, fix $\varepsilon > 0$ and let $A_i' = (1 - \varepsilon)A_i$ and $A_i'' = (1 + \varepsilon)A_i$ in the notation of Lemma 5.9.19. By that lemma, for N sufficiently large, we have

$$T^{c_i}(A_i') \subseteq \phi(A_i) \subseteq T^{c_i}(A_i'').$$

Thus,

$$m(T^{c_i}(A_i')) \leq m_*(\phi(A_i)) \leq m^*(\phi(A_i)) \leq m(T^{c_i}(A_i'')).$$

Using the previous lemma, we have

$$(1-\varepsilon)^n|J(c_i)|m(A_i) \le m_*(\phi(A_i)) \le m^*(\phi(A_i)) \le (1+\varepsilon)^n|J(c_i)|m(A_i).$$

Summing over i and letting $\varepsilon \to 0$, we see that $m_*(B) = m^*(B)$, so B is measurable by Exercise 6.2.20.

Using the continuity of $J(x)$, we see that for N large enough we have

$$(1-\varepsilon)\sum_{i=1}^{N}|J(c_i)|m(A_i) \le \int_A |J(x)|\,dx \le (1+\varepsilon)\sum_{i=1}^{N}|J(c_i)|m(A_i).$$

Since $m(B)$ is bounded above and below by the same quantities, letting $\varepsilon \to 0$, we may conclude that

$$m(B) = \int_A |J(x)|\,dx.$$

LEMMA 6.7.4. *With notation as in Theorem* 6.7.1*, if A is bounded, then B is measurable, and we have*

$$m(B) = \int_A |J(x)|\,dx.$$

Proof. If A is a disjoint union of rectangles, then the result follows from the previous lemma. In general, let R be a rectangle containing A, and let $\varepsilon > 0$. Then, by the definition of inner and outer Lebesgue measure, there exist sets A_1 and A_2 with $A_1 \subseteq A \subseteq A_2$ such that A_2 is a disjoint union of rectangles, $R \setminus A_1$ is a disjoint union of rectangles, and $m(A_2 \setminus A_1) < \varepsilon$.

Then

$$m^*(B) \le m(\phi(A_2)) = \int_{A_2} |J(x)|\,dx$$

and

$$m_*(B) \ge m(\phi(A_1)) = \int_{A_1} |J(x)|\,dx,$$

and the result follows from Proposition 6.4.26.

We now have the change of variables formula for the characteristic function of a bounded measurable set.

Proof of Theorem 6.7.1. If f is a measurable simple function of bounded support, then the result follows from the previous lemma. Now let f be a bounded measurable function on a bounded measurable set A, and let $\varepsilon > 0$. By Exercise 6.4.18 there exist measurable simple functions f_1 and f_2 such that $f_1 \le f \le f_2$ and

$$\int_A (f_2 - f_1)\,dm < \varepsilon.$$

By change of variables for simple functions, we have

$$\int_A (f_2(\phi(x)) - f_1(\phi(x)))|J(x)|\,dx < \varepsilon,$$

so that

$$\int_A (f(\phi(x)) - f_1(\phi(x)))|J(x)|\,dx < \varepsilon.$$

Thus,

$$\begin{aligned}\left|\int_A f(\phi(x))|J(x)|\,dx - \int_B f(y)\,dy\right| &\leq \left|\int_A (f(\phi(x)) - f_1(\phi(x))|J(x)|\,dx\right| \\ &\quad + \left|\int_A f_1(\phi(x))|J(x)|\,dx - \int_B f_1(y)\,dy\right| \\ &\quad + \left|\int_B (f(y) - f_1(y))\,dy\right|.\end{aligned}$$

The middle term is equal to 0 by change of variables for simple functions, and the first and last terms are each less than ε. It follows that

$$\int_A f(\phi(x))|J(x)|\,dx = \int_B f(y)\,dy,$$

which establishes the change of variables formula for a bounded measurable function on a bounded measurable set.

EXERCISE 6.7.5. Finish the proof by extending the change of variables formula to general measurable functions on (possibly unbounded) general measurable sets in $\mathbb{R}^n$.

6.8. Independent Projects

6.8.1. The Haar Integral: Construction.

In the following, G will be a locally compact separable topological group, unless otherwise specified.

DEFINITION 6.8.1. Let μ be a σ-finite regular Borel measure on G. We say that μ is a *left Haar measure* on G if the following properties hold:

(a) For any Borel set $E \subseteq G$ and any $x \in G$, we have $\mu(xE) = \mu(E)$ (that is, μ is *invariant under left translation*).
(b) If C is compact, then $\mu(C) < \infty$.
(c) If U is a nonempty open set, then $\mu(U) > 0$.

REMARK 6.8.2. To define a right Haar measure, we replace the condition (a) with the following:

(a′) $\mu(Ex) = \mu(E)$.

THEOREM 6.8.3. *If G is a locally compact group, then there is a left Haar measure on G. This measure is unique up to multiplication by a positive constant.*

There is a similar statement about right Haar measure.

EXERCISE 6.8.4. Let μ be a left Haar measure on G. For any measurable set E, define $\nu(E) = \mu(E^{-1})$. Show that ν is a right Haar measure on G.

DEFINITION 6.8.5. Let $\mathcal{C}$ be a collection of subsets of G. Then $\mathcal{C}$ is a *sequential covering class* if $\varnothing \in \mathcal{C}$ and, for every $A \subseteq G$, there exists a countable subcollection $\{E_n\} \subseteq \mathcal{C}$ such that

$$A \subseteq \bigcup_{n=1}^{\infty} E_n.$$

Given a sequential covering class $\mathcal{C}$ and a nonnegative function $\tau : \mathcal{C} \longrightarrow \mathbb{R} \cup \{\infty\}$ such that $\tau(\varnothing) = 0$, we can construct an outer measure μ^* on G by defining

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \tau(E_n) \middle| E_n \in \mathcal{C},\ A \subseteq \bigcup_{n=1}^{\infty} E_n \right\}$$

for any $A \subseteq G$.

EXERCISE 6.8.6. Show that μ^* is an outer measure.

Using the general construction in this chapter (see Definition 6.2.10), we obtain from μ^* a class of measurable sets and a measure μ on G. Thus, to construct a Haar measure, we need only to find an acceptable sequential covering class $\mathcal{C}$ and a suitable function τ satisfying the required properties.

Let $\mathcal{C}$ consist of G, $\varnothing$, and the class $\mathcal{C}'$ of all nonempty open sets whose closures are compact. Since $G \in \mathcal{C}$, the collection $\mathcal{C}$ is a sequential covering class. Since G is locally compact, every compact set is covered by a finite subcollection of $\mathcal{C}'$. Now, we must construct τ. The construction that follows assumes that the topology on G is induced by a metric. If G does not have a metric topology, other techniques are necessary to construct a suitable function τ.

Let $S \subseteq G$ be an open set. Note that, for all $U \in \mathcal{C}'$, the set

$$\{xS \mid x \in \overline{U}S^{-1}\}$$

is an open covering of $\overline{U}$. Since $\overline{U}$ is compact, there is a finite subcollection of this set that covers $\overline{U}$. Let $(U : S)$ be the smallest number of left translations of S that will cover $\overline{U}$.

EXERCISE 6.8.7. Let $A, B \in \mathcal{C}'$ and let $C \subseteq G$ be an open set. Show that

$$1 \leq (A : C) \leq (A : B)(B : C).$$

Now, we fix an element $U_0 \in \mathcal{C}'$. For every open set $S \subseteq G$ and every $U \in \mathcal{C}'$, we define

$$\tau_S(U) = \frac{(U : S)}{(U_0 : S)}.$$

It follows from Exercise 6.8.7 that

$$\frac{1}{(U_0 : U)} \leq \tau_S(U) \leq (U : U_0),$$

so $\tau_S(U)$ is a bounded function of S.

Now let $V_n = B_{1/n}(e)$, where e is the identity in G. For any $U \in \mathcal{C}'$, the sequence $\tau_{V_n}(U)$ is bounded. So, by the generalized limit of Section 4.5, we may define

$$\tau(U) = \mathrm{L.\,i.\,m.}_n\, \tau_{V_n}(U).$$

We also set $\tau(\varnothing) = 0$, and if G is not compact, we set $\tau(G) = \infty$.

THEOREM 6.8.8.

(i) *For all $U \in \mathcal{C}'$, we have $0 < \tau(U) < \infty$.*
(ii) *If $U_1, U_2 \in \mathcal{C}'$, then $\tau(U_1 \cup U_2) \leq \tau(U_1) + \tau(U_2)$.*
(iii) *If $U_1, U_2 \in \mathcal{C}'$ and $d(U_1, U_2) > 0$, then $\tau(U_1 \cup U_2) = \tau(U_1) + \tau(U_2)$.*
(iv) *If $U_1, U_2 \in \mathcal{C}'$ and $U_1 \subseteq U_2$, then $\tau(U_1) \leq \tau(U_2)$.*

EXERCISE 6.8.9. Prove this theorem.

Let μ denote the measure associated to the outer measure μ^* constructed from this τ as above.

EXERCISE 6.8.10. Show that μ is a left Haar measure on G.

A standard example of Haar measure is Lebesgue measure on $(\mathbb{R}, +)$. Because $(\mathbb{R}, +)$ is abelian, this measure is both left and right invariant (earlier in the chapter we had denoted this measure by $dm(x)$ and then, later, simply by dx). This measure is ordinarily normalized so that the measure of $[0, 1]$ is equal to 1. That is,

$$\int_{[0,1]} dx = 1.$$

In Fourier analysis on $\mathbb{R}$, there will be a natural reason to consider the measure of $[0, 1]$ to be $1/\sqrt{2\pi}$. We have

$$\int_{\mathbb{R}} f(x + y)dx = \int_{\mathbb{R}} f(x)dx.$$

For Haar measure on the multiplicative group $\mathbb{R}_+^\times$ of positive reals, we can take $d^\times x = dx/x$. This is a consequence of the fact that, for the additive Haar measure dx, we have

$$\int_{\mathbb{R}} f(ax)dx = a \int_{\mathbb{R}} f(x)dx$$

for $a \in \mathbb{R}_+^\times$. This follows from the definition of the area of a rectangle. In fact, for any $a \in \mathbb{R}^\times$, we have

$$\int_{\mathbb{R}} f(ax)dx = |a| \int_{\mathbb{R}} f(x)dx.$$

Thus we can extend $d^\times x$ to all of $\mathbb{R}^\times$ by setting $d^\times x = dx/|x|$. We have

$$\int_{\mathbb{R}^\times} f(ax) d^\times x = \int_{\mathbb{R}^\times} f(ax) \frac{dx}{|x|} = \int_{\mathbb{R}^\times} f(x) \frac{dx}{|x|} = \int_{\mathbb{R}^\times} f(x) d^\times x.$$

This, of course, is a very special case of the change of variables formula. Observe that, with this Haar measure on $\mathbb{R}^\times$, the measure of $[1, e]$ is equal to 1.

From now on, we will write $d_r x$ for the right Haar measure on a group and $d_l x$ for the left Haar measure on a group. Note that for $f \in C_c(G)$, we have

$$\int_G f(xy) d_r x = \int_G f(x) d_r x$$

and

$$\int_G f(yx) d_l x = \int_G f(x) d_l x$$

for any $y \in G$.

6.8.2. The Haar Integral: Facts and Examples.

Not every right Haar measure is also a left Haar measure.

EXAMPLE 6.8.11 **($ax + b$ group)**. The "$ax + b$ group" is the subgroup of $GL_2(\mathbb{R})$ consisting of matrices of the form $\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$, where $a \in \mathbb{R}^\times_+$ and $b \in \mathbb{R}$. This group acts on the real line with the matrix $\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$ mapping the point x to the point $ax + b$. In other words, this is the group of affine transformations of the real line.

Let db be the Lebesgue measure on $\mathbb{R}$, and let $d^\times a = \frac{da}{a}$ be the Haar measure on $\mathbb{R}^\times_+$. Then, $d^\times a\, db$ is a right Haar measure on G, and $a^{-1} d^\times a\, db$ is a left Haar measure on G. For $f \in C_c(G)$,

$$\int_G f\left[\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}\begin{pmatrix} c & d \\ 0 & 1 \end{pmatrix}\right] d^\times a\, db = \int_G f\left[\begin{pmatrix} ac & ad+b \\ 0 & 1 \end{pmatrix}\right] d^\times a\, db$$
$$= \int_G f\left[\begin{pmatrix} ac & b \\ 0 & 1 \end{pmatrix}\right] d^\times a\, db = \int_G f\left[\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}\right] d^\times a\, db.$$

Notice that this demonstration uses the fact that db is invariant under addition and $d^\times a$ is invariant under multiplication.

EXERCISE 6.8.12. Show that, for $f \in C_c(G)$,

$$\int_G f\left[\begin{pmatrix} c & d \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}\right] a^{-1} d^\times a\, db = \int_G f\left[\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}\right] a^{-1} d^\times a\, db.$$

EXERCISE 6.8.13. Find a continuous complex-valued function on the $ax + b$ group that is integrable with respect to left Haar measure but is not integrable with respect to right Haar measure.

DEFINITION 6.8.14. If every left Haar measure on a particular group G is also a right Haar measure, then we say that G is *unimodular*.

Suppose that $y \in G$. Then the map

$$f \longmapsto \int_G f(yx) d_r x$$

also defines a right Haar integral on G. That is, $\int_G f(yxz) d_r x = \int_G f(yx) d_r x$ for all $f \in C_c(G)$. Since the right Haar measure is unique up to multiplication by a positive constant, there exists $\Delta(y) \in \mathbb{R}_+^\times$ such that

$$\int_G f(yx) d_r x = \Delta(y) \int_G f(x) d_r x.$$

Note that, in particular, G is unimodular if and only if Δ is identically equal to 1.

EXERCISE 6.8.15. Show directly from the definition that Δ is a homomorphism from G into the group $\mathbb{R}_+^\times$.

PROPOSITION 6.8.16. *The map Δ is a continuous homomorphism.*

Proof. Since Δ is a homomorphism, we need only show continuity at the identity e of G. Let $f \in C_c(G)$ with support contained in the compact set C. Note that

$$\begin{aligned}
\left| (\Delta(y) - 1) \int_G f(x) d_r x \right| &= \left| \Delta(y) \int_G f(x) d_r x - \int_G f(x) d_r x \right| \\
&= \left| \int_G f(yx) d_r x - \int_G f(x) d_r x \right| \\
&= \left| \int_G (f(yx) - f(x)) d_r x \right| \\
&\leq \int_G |f(yx) - f(x)| d_r x.
\end{aligned}$$

EXERCISE 6.8.17. Show that, for any $\varepsilon > 0$, there exists a compact neighborhood $\mathcal{U}$ of e such that

$$|f(yx) - f(x)| < \varepsilon$$

for all $y \in \mathcal{U}$ and $x \in C \cup \mathcal{U}^{-1}C$. Now complete the proof.

EXERCISE 6.8.18. Show that the following classes consist of unimodular groups:

(i) abelian groups,
(ii) compact groups,
(iii) discrete groups.

EXERCISE 6.8.19. Suppose that G is a locally compact group and every normal subgroup of G lies in the center of G. Show that G is unimodular.

EXERCISE 6.8.20. Show that

$$\int_G f(x)\Delta(x)d_r x = \int_G f(yx)\Delta(x)d_r x,$$

so $d_l x = \Delta(x)d_r x$ is a left Haar measure on G.

EXERCISE 6.8.21. If $G_1, G_2, \ldots, G_n$ are locally compact groups with right Haar measures $d_r x_1, d_r x_2, \ldots, d_r x_n$, respectively, then the product measure $d_r x_1 \cdot d_r x_2 \cdot \cdots \cdot d_r x_n$ is a right Haar measure for $G = G_1 \times G_2 \times \cdots \times G_n$, and G is unimodular if and only if all of the G_i's are.

EXAMPLE 6.8.22. On $\mathbb{R}^n$, $dx_1 dx_2 \cdots dx_n$ gives the usual Lebesgue measure as a Haar measure.

EXAMPLE 6.8.23 **(Haar measure for $GL_n(\mathbb{R})$)**. For $x \in M_n(\mathbb{R})$, we can write $x = (x_1, \ldots, x_n)$, where x_j is the column vector $\begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}$. For any $u \in GL_n(\mathbb{R})$, ux can now be written as $ux = (ux_1, \ldots, ux_n)$. By the change of variables theorem, for $f \in C_c(M_n(\mathbb{R}))$,

$$\begin{aligned}\int_{M_n(\mathbb{R})} f(ux)dx &= \int_{M_n(\mathbb{R})} f(ux_1, ux_2, \ldots, ux_n)dx_1 dx_2 \cdots dx_n \\ &= |\det(u)|^{-1} \int_{M_n(\mathbb{R})} f(x_1, ux_2, \ldots, ux_n)dx_1 dx_2 \cdots dx_n,\end{aligned}$$

so repeating this step, we have

$$\int_{M_n(\mathbb{R})} f(ux)dx = |\det(u)|^{-n} \int_{M_n(\mathbb{R})} f(x_1, x_2, \ldots, x_n)dx_1\ dx_2 \cdots dx_n.$$

EXERCISE 6.8.24. Using this computation, show that $d^\times x = |\det(x)|^{-n}dx$ is a left Haar measure on $GL_n(\mathbb{R})$. Show that the same expression also defines a right Haar measure on $GL_n(\mathbb{R})$.

Observe that the Haar measure on $GL_n(\mathbb{R})$ generalizes the Haar measure $d^\times x$ on $GL_1(\mathbb{R}) = \mathbb{R}^\times$.

6.8.3. Representations of Locally Compact Groups.

In Chapter 7 we consider Fourier analysis on locally compact abelian groups, where the irreducible representations are one-dimensional. For compact groups, irreducible unitary representations are all finite-dimensional. The big difference between the abelian and compact cases and the general *locally compact* case is the appearance of infinite-dimensional irreducible unitary representations.

In this project, we will always assume that the topology on the locally compact group is separable.

DEFINITION 6.8.25. Let G be a locally compact group and let V be a Hilbert space. A *representation* of G on V is a map

$$T : G \longrightarrow \mathcal{BL}(V)$$

such that

(a) $T(gh) = T(g)T(h)$ for all $g, h \in G$,
(b) $T(e) = I$, where $e \in G$ is the identity element and $I \in \mathcal{BL}(V)$ is the identity operator, and
(c) T is strongly continuous, that is, for each $v \in V$ the map

$$g \longmapsto T(g)v$$

is continuous.

We shall often write (T, V) for such a representation. If V_0 is a subspace of V, we say that V_0 is *invariant* under T if $T(g)V_0 \subseteq V_0$ for all $g \in G$.

EXERCISE 6.8.26. Show that, for any representation (T, V), the subspaces $\{0\}$ and V are invariant.

A representation (T, V) is *irreducible* if the only closed subspaces of V that are invariant under T are $\{0\}$ and V.

If (T, V) is a representation of a locally compact group G and V is an infinite-dimensional space, we say that T is an *infinite-dimensional* representation of G. On the other hand, if V is finite-dimensional, we say that T is *finite-dimensional*, and we refer to the dimension of V as the *degree* of T.

A representation (T, V) is a *unitary* representation if $T(g)$ is a unitary operator on V for all $g \in G$. In this case, we may say that T is a homomorphism from G to $\mathcal{U}(V)$ satisfying the strong continuity condition above.

EXERCISE 6.8.27. We say that a representation of G on V is *weakly continuous* if, for each pair $v, w \in V$, the map

$$g \longmapsto (T(g)v|w)$$

is continuous from G to $\mathbb{C}$. Prove that, for unitary representations, weak and strong continuity are equivalent.

EXERCISE 6.8.28. Let (T, V) be a unitary representation of a locally compact group G. Show that if $V_0 \subseteq V$ is a closed subspace invariant under T, then $V_0^{\perp}$ is also a closed subspace invariant under T.

DEFINITION 6.8.29. Let (T, V) be a unitary representation of a locally compact group G. If $V_0 \subseteq V$ is a closed subspace invariant under T, then (T, V_0) is a unitary representation of G on V_0 that is called a *subrepresentation* of T.

In some situations a given representation of a group G can be decomposed into its subrepresentations. To make this notion precise, we require the following definition.

DEFINITION 6.8.30. Let G be a locally compact group and let (T_1, V_1) and (T_2, V_2) be two unitary representations of G. The *direct sum* of T_1 and T_2 is the representation $T_1 \oplus T_2$ of G on the Hilbert space direct sum $V_1 \oplus V_2$ defined by

$$[(T_1 \oplus T_2)(g)](v_1, v_2) = (T_1(g)v_1, T_2(g)v_2).$$

EXERCISE 6.8.31. Show that this defines a unitary representation of G.

At this point, we observe that a representation of a locally compact group G involves a Hilbert space that *a priori* has nothing to do with G. If we have two such unitary representations (T_1, V_1) and (T_2, V_2), we may ask if they are the same representation in some sense. This leads to the notion of unitary equivalence.

DEFINITION 6.8.32. We say that (T_1, V_1) and (T_2, V_2) are *unitarily equivalent* if there exists a unitary operator $A : V_1 \longrightarrow V_2$ such that

$$AT_1(g) = T_2(g)A \quad \text{for all } g \in G.$$

Such an operator A is called an *intertwining operator* for T_1 and T_2.

EXERCISE 6.8.33. Show that unitary equivalence is an equivalence relation.

The collection of equivalence classes of unitary representations of a locally compact group G is called the *unitary dual* of G and is denoted by $\hat{G}$.

EXERCISE 6.8.34. Let (T, V) be a unitary representation of a locally compact group G and let $V_0 \subseteq V$ be a closed subspace invariant under T. Show that the isometric isomorphism

$$\begin{array}{rcl} V_0 \oplus V_0^{\perp} & \longrightarrow & V \\ (v, v') & \longmapsto & v + v' \end{array}$$

is a unitary equivalence of representations of G.

EXERCISE 6.8.35. Let $\{V_i\}_{i \in \mathbb{N}}$ be a countable collection of Hilbert spaces. Show that

$$\bigoplus_{i \in \mathbb{N}} V_i = \left\{ (v_i) \;\middle|\; \sum_{i \in \mathbb{N}} \|v_i\|^2 < \infty \right\}$$

is a Hilbert space with the obvious inner product.

DEFINITION 6.8.36. A unitary representation (T, V) of a locally compact group G is *completely reducible* if T can be written as a (possibly countable) direct sum of irreducible unitary representations.

EXERCISE 6.8.37. Show that a finite-dimensional unitary representation of a locally compact group G is completely reducible.

The irreducible unitary representations of a locally compact group are the basic building blocks of harmonic analysis on the group. We need a criterion to determine irreducibility of a given representation. The following theorem is extremely useful in this pursuit.

THEOREM 6.8.38. *Let (T, V) be a unitary representation of a locally compact group G. Then T is irreducible if and only if every $A \in \mathcal{BL}(V)$ that satisfies the condition $AT(g) = T(g)A$ for all $g \in G$ is of the form $A = cI$ for some $c \in \mathbb{C}$.*

One-half of this theorem is easy to prove.

EXERCISE 6.8.39. If T is not irreducible (*reducible*), there is a closed invariant subspace $V_0 \subseteq V$ such that $V_0 \neq \{0\}, V$. Show that the orthogonal projection of V onto V_0 is an intertwining operator for the representation T.

The other half of the proof relies on the spectral theorem for noncompact selfadjoint operators.

Let G be a locally compact group and assume that G is unimodular. The most important unitary representations of G are the left and right regular representations. Let V denote the Hilbert space $L^2(G)$. If G were not unimodular, then the spaces $L^2(G, d_l x)$ and $L^2(G, d_r x)$ would be distinct.

EXERCISE 6.8.40. Let G be a locally compact group that is not unimodular. Find a function in $L^2(G, d_l x)$ that is not in $L^2(G, d_r x)$ and a function in $L^2(G, d_r x)$ that is not in $L^2(G, d_l x)$.

DEFINITION 6.8.41. Let G be a unimodular group. The *right regular representation* R of G on $V = L^2(G)$ is defined by

$$[R(x)f](y) = f(yx)$$

and the *left regular representation* L of G on V is defined by

$$[L(x)f](y) = f(x^{-1}y).$$

EXERCISE 6.8.42. Show that R and L are unitary representations of G on $V = L^2(G)$.

EXERCISE 6.8.43. Show that R and L are unitarily equivalent via the intertwining operator A defined by

$$[Af](x) = f(x^{-1}).$$

EXERCISE 6.8.44. Let $G = (\mathbb{R}, +)$ and let $(R, L^2(\mathbb{R}))$ be the right regular representation of G. Show that the space $C_c^\infty(\mathbb{R})$ is invariant under R.

REMARK 6.8.45. Note that $C_c^\infty(\mathbb{R})$ is not a closed subspace of $L^2(\mathbb{R})$.

Chapter 7

Fourier Analysis on Locally Compact Abelian Groups

> [The classical theory of Fourier series] has been a source of new ideas for analysts during the last two centuries, and is likely to be so in years to come. Many basic notions and results of the theory of functions have been obtained by mathematicians while working on trigonometric series. Conceivably these discoveries might have been made in different contexts, but in fact they came to life in connexion with the theory of trigonometric series. It was not accidental that the notion of function generally accepted now was first formulated in the celebrated memoir of Dirichlet (1837) dealing with the convergence of Fourier series; or that the definition of Riemann's integral in its general form appeared in Riemann's Habilitationsschrift devoted to trigonometric series; or that the theory of sets, one of the most important developments of nineteenth-century mathematics, was created by Cantor in his attempts to solve the problem of the sets of uniqueness for trigonometric series. In more recent times, the integral of Lebesgue was developed in close connexion with the theory of Fourier series, and the theory of generalized functions (distributions) with that of Fourier integrals.
>
> – *Antoni Zygmund*
> *Trigonometric Series*

Challenge Problems: VII

PROBLEM 7.1. Find an example of a nontrivial (that is, other than $\{0\}$ or $L^2(\mathbb{R})$) closed invariant subspace of $L^2(\mathbb{R})$ under the right regular representation R.

PROBLEM 7.2. Let $f \in L^1(\mathbb{R})$ and $g \in L^p(\mathbb{R})$ for $1 < p < \infty$ and define $T_f(g) = f * g$. Find the norm of T_f as a bounded linear operator on $L^p(\mathbb{R})$.

PROBLEM 7.3. Let $T : L^2(\mathbb{R}) \longrightarrow L^2(\mathbb{R})$ be a linear map which is norm-preserving on characteristic functions. Show that T can be extended to an isometry from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$. Show that this is false if $L^2(\mathbb{R})$ is replaced by $L^2(\mathbb{R}^2)$.

PROBLEM 7.4. Show that there exists a Borel set $E \subseteq \mathbb{R}$ such that $0 < m(E \cap I) < m(I)$ for all intervals $I \subseteq \mathbb{R}$.

PROBLEM 7.5. Let A be a unitary operator on $L^2(\mathbb{R})$ that commutes with translation. Give a precise description of A.

The study of trigonometric series began over two hundred years ago (before the birth of Fourier) in an attempt to find solutions to certain partial differential equations, in particular, the wave equation, which describes the motion of a vibrating string. Later, Fourier himself employed trigonometric series to solve the equation for the diffusion of heat in finite bodies, that is, the heat equation. The basic idea involved in these solutions was that of "superposition" (i.e., addition) of the elementary trigonometric functions $\sin(\frac{\pi kx}{L})$, $\cos(\frac{\pi kx}{L})$, $k = 0, 1, 2, 3, \ldots$, and their products. In this same era, the Fourier and Laplace transforms were used (by Fourier and Laplace) in the solution of the heat equation for infinite continuous bodies.

Naturally, many problems arose relating to the convergence of the resulting series and integrals, and a number of mathematicians in the early nineteenth century studied these questions. The theory of Fourier (and trigonometric) series and integrals grew in mathematical content during the nineteenth century as real and complex variable techniques became more sophisticated, and, for over two centuries, the subject has been an active branch of pure mathematics.

It is not our intention to trace the development of these ideas (for this, see Hawkins [**8**]). We simply note that Fourier analysis had its beginnings in the theory of partial differential equations and that the roots of harmonic analysis on locally compact groups lie, in part, in soil tilled two centuries ago.

7.1. Fourier Analysis on the Circle

As usual, let $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$, the unit circle in the complex plane. As pointed out in Chapter 4, $\mathbb{T}$ is a compact abelian group. In analysis, $\mathbb{T}$ plays a dual role. On the one hand, $\mathbb{T}$ is topologically isomorphic to S^1, the unit circle in $\mathbb{R}^2$, and on the other hand, $\mathbb{T}$ is the 1-torus in $\mathbb{C}$. The generalization of $\mathbb{T}$ thus goes in two directions, first to S^{n-1} in $\mathbb{R}^n$ and second to the n-torus $\mathbb{T}^n = \mathbb{T} \times \mathbb{T} \times \cdots \times \mathbb{T}$ in $\mathbb{C}^n$. In the first case, S^{n-1} is the domain of the study of spherical harmonics, and in the second case, $\mathbb{T}^n$ is the domain of the study of multiple Fourier series.

DEFINITION 7.1.1. For $n \in \mathbb{Z}$, define $\chi_n : \mathbb{T} \longrightarrow \mathbb{T}$ by $\chi_n(e^{i\theta}) = e^{in\theta} \in \mathbb{T}$.

Note that, for each n, χ_n is a continuous homomorphism from $\mathbb{T}$ to $\mathbb{T}$. These maps are called the *unitary characters* (or basic harmonics) on $\mathbb{T}$. Their general role in Fourier analysis will be discussed in the next section.

DEFINITION 7.1.2. If $f \in L^1(\mathbb{T})$ and $n \in \mathbb{Z}$, we define

$$\hat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{i\theta}) e^{-in\theta} d\theta.$$

The complex number $\hat{f}(n)$ is called the n-th *Fourier coefficient* of f, and $\hat{f}$ can be regarded as a map from $\mathbb{Z}$ to $\mathbb{C}$.

Since $\chi_n(e^{i\theta}) = \cos(n\theta) + i \sin(n\theta)$, we are writing Fourier coefficients in terms of the classical trigonometric functions.

DEFINITION 7.1.3. Let $f \in L^1(\mathbb{T})$. The *Fourier series* of f is the series

$$\sum_{n=-\infty}^{\infty} \hat{f}(n) e^{in\theta}.$$

At this point, we have no idea whether this series converges, how it converges, and, if it converges, whether it converges to f in any sense whatsoever. Before beginning the study of the convergence of Fourier series, we discuss several important properties of Fourier coefficients.

FACTS 7.1.4. (1) Riemann-Lebesgue lemma: If $f \in L^1(\mathbb{T})$, then

$$\lim_{|n| \to \infty} \hat{f}(n) = 0.$$

Proof. It suffices to prove this for characteristic functions of subintervals of $(-\pi, \pi]$. The proof is left to the reader.

If f is a complex-valued function on $\mathbb{T}$, we will often write $f(\theta)$ for $f(e^{i\theta})$. This will simplify the notation somewhat.

(2) If $f, g \in L^1(\mathbb{T})$, we define the *convolution* of f and g by

$$(f * g)(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi) g(\theta - \phi)\, d\phi.$$

If $f, g \in L^1(\mathbb{T})$, then $f * g \in L^1(\mathbb{T})$.

Proof. This follows from Fubini's theorem.

Hence $L^1(\mathbb{T})$ is an algebra over $\mathbb{C}$ with multiplication defined by convolution.

EXERCISE 7.1.5. (a) If $f, g \in L^1(\mathbb{T})$, show that $f * g = g * f$.
(b) Show that $L^1(\mathbb{T})$ does not have a multiplicative identity.

(3) The space $C^\infty(\mathbb{T})$ is an algebra under convolution; that is, if $f, g \in C^\infty(\mathbb{T})$, then $f * g \in C^\infty(\mathbb{T})$.

Proof. Differentiate under the integral sign.

(4) If $f \in C^\infty(\mathbb{T})$, then $\lim_{|n|\to\infty} n^k \hat{f}(n) = 0$ for all $k \in \mathbb{N}$. That is, the Fourier coefficients of f *decrease rapidly at infinity.*

Proof. Integrate by parts.

(5) For $f, g \in L^1(\mathbb{T})$, $\widehat{(f * g)}(n) = \hat{f}(n)\hat{g}(n)$.

Proof. Use Fubini's theorem and the translation invariance of Haar measure.

(6) If f is differentiable on $\mathbb{T}$, then

$$in\widehat{f'} = \hat{f}$$

for all $n \in \mathbb{Z}$.

Proof. Integrate by parts.

(7) Note that the characters χ_n are continuous functions and hence are in $L^2(\mathbb{T})$. These characters $(\chi_n)_{n\in\mathbb{Z}}$ form an orthonormal set in $L^2(\mathbb{T})$. That is,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \chi_n(\theta)\overline{\chi_m(\theta)}\,d\theta = \begin{cases} 1 & n = m, \\ 0 & n \neq m. \end{cases}$$

(8) For $N \in \mathbb{N}$, define

$$S_N(f)(\theta) = \sum_{n=-N}^{N} \hat{f}(n)e^{in\theta}$$

for $f \in L^1(\mathbb{T})$. If $1 \leq p \leq \infty$, then $L^p(\mathbb{T}) \subseteq L^1(\mathbb{T})$, and $S_N(f) \in L^p(\mathbb{T})$.

EXERCISE 7.1.6 **(Bessel's inequality)**. Show that if $f \in L^2(\mathbb{T})$, then

$$\|f\|_2^2 \geq \|S_N(f)\|_2^2 = \sum_{n=-N}^{N} |\hat{f}(n)|^2.$$

Conclude that $\sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2$ converges, so that the sequence $(\widehat{f}(n))_{n\in\mathbb{Z}} \in L^2(\mathbb{Z})$.

EXERCISE 7.1.7. For $f \in L^1(\mathbb{T})$, define $\tilde{f}(\theta) = \overline{f(-\theta)}$. Show that $\tilde{f} \in L^1(\mathbb{T})$ and $\widehat{(\tilde{f})}(n) = \overline{\hat{f}(n)}$.

EXAMPLE 7.1.8. Let $f(\theta) = \theta$ for $-\pi < \theta \leq \pi$. Then

$$\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \theta e^{-in\theta}\,d\theta = \frac{1}{2\pi}\left(\frac{\theta e^{-int}}{-in}\bigg|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} \frac{e^{-in\theta}}{-in}\,d\theta\right) = \frac{i(-1)^n}{n}.$$

EXERCISE 7.1.9. Explain why the Fourier series above does not decrease rapidly at infinity.

EXERCISE 7.1.10. Compute the Fourier series of $f(\theta) = \theta^2$ for $-\pi < \theta \leq \pi$.

In considering the convergence of the Fourier series of a function f in $L^1(\mathbb{T})$, we must make a decision about computing partial sums. On the one hand, we can consider symmetric partial sums. Thus, if $N \in \mathbb{N}$, we may consider the sum $S_N(f)(\theta) = \sum_{n=-N}^{N} \hat{f}(n)e^{in\theta}$. Of course, we could also consider, for $M, N \in \mathbb{N}$, partial sums of the form $S_{M,N}(f)(\theta) = \sum_{n=-M}^{N} \hat{f}(n)e^{in\theta}$. The convergence of the Fourier series of f will in many cases depend on which form of the partial sums we choose. However, if $\sum_{n=-\infty}^{\infty} |\hat{f}(n)|$ converges, then the convergence properties of the Fourier series of f do not depend on which of the above methods of partial summation we choose.

EXERCISE 7.1.11. Consider $f(\theta) = \theta$ for $-\pi < \theta \leq \pi$. Compare the results of both methods of summation.

EXERCISE 7.1.12. Let $f \in C^2(\mathbb{T})$. Show that the Fourier series of f converges absolutely, that is, $\sum_{n=-\infty}^{\infty} |\widehat{f}(n)|$ converges.

EXERCISE 7.1.13. Compute the Fourier series of the following functions on $\mathbb{T}$, where $-\pi < \theta \leq \pi$.

(i) $f(\theta) = |\theta|$.
(ii) $f(\theta) = e^{\theta}$.
(iii) $f(\theta) = \theta \cos\theta$.
(iv) $f(\theta) = \cos(\alpha\theta)$ for $\alpha \notin \mathbb{Z}$.
(v) Suppose $[a, b] \subseteq (-\pi, \pi]$. Compute the Fourier series of the characteristic function of $[a, b]$.

7.1.1. The C^1-Theory.

DEFINITION 7.1.14. Let $f : \mathbb{T} \longrightarrow \mathbb{C}$. We say that f is *piecewise continuous* if f is continuous except perhaps at a finite number of points $p_1, \ldots, p_k$ and f has left and right limits at these points. We say that f is *piecewise smooth* if f is continuously differentiable except perhaps at a finite number of points at which f has both a left- and right-hand derivative.

In this section, we investigate the convergence of Fourier series of piecewise smooth functions. We first look at the symmetric partial sums of an L^1 function f. Since the sum S_N is finite,

$$\begin{aligned} S_N(f)(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi) \sum_{n=-N}^{N} e^{-in\phi} e^{in\theta} \, d\phi \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\phi) \frac{\sin(N + \frac{1}{2})(\theta - \phi)}{\sin \frac{1}{2}(\theta - \phi)} \, d\phi. \end{aligned}$$

DEFINITION 7.1.15. The *Dirichlet kernel* is defined by

$$D_N(\theta) = \sum_{i=-N}^{N} e^{in\theta}.$$

Observe that $S_N(f) = f * D_N$.

EXERCISE 7.1.16.

(i) Show that $D_N(0) = 2N + 1$ and

$$D_N(\theta) = \frac{\sin(N + \frac{1}{2})\theta}{\sin \frac{1}{2}\theta}$$

for $\theta \in (-\pi, \pi] \setminus \{0\}$.

(ii) Show that $\int_{-\pi}^{\pi} D_N(\theta)\, d\theta = 2\pi$.

It is now clear that the convergence of the Fourier series of f is closely related to the properties of the Dirichlet kernel. The following lemma will be useful to help us analyze the Dirichlet kernel and will also aid us in the study of the Fourier transform on $\mathbb{R}$.

LEMMA 7.1.17. *Let $g \in L^1([a, b])$, for $[a, b] \subseteq \mathbb{R}$. Then*

$$\lim_{\lambda \to \infty} \int_a^b g(t) \sin(\lambda t)\, dt = 0.$$

Proof. We can approximate g by a step function in the L^1 norm and integrate.

COROLLARY 7.1.18. *Let $g \in L^1(\mathbb{T})$. Then $\lim_{\lambda \to \infty} \int_{-\pi}^{\pi} g(\theta) \sin(\lambda\theta)\, d\theta = 0$.*

COROLLARY 7.1.19. *Let $g \in L^1(\mathbb{R})$. Then $\lim_{\lambda \to \infty} \int_{\mathbb{R}} g(t) \sin(\lambda t)\, dt = 0$.*

Proof. Given $\varepsilon > 0$, choose $A > 0$ such that

$$\int_{|t| \geq A} |g(t)|\, dt < \frac{\varepsilon}{2}.$$

Then choose L such that for $\lambda > L$ we have

$$\left| \int_{-A}^{A} g(t) \sin(\lambda t)\, dt \right| < \frac{\varepsilon}{2}.$$

We are now ready to prove the first basic theorem on pointwise convergence of Fourier series.

THEOREM 7.1.20. *Let f be a piecewise smooth function on $\mathbb{T}$. If f is continuous at $e^{i\theta} \in \mathbb{T}$, then $\lim_{N \to \infty} S_N(\theta) = f(\theta)$.*

Proof. For $\phi \in (-\pi, \pi]$ define $g(\phi) = \frac{f(\theta-\phi)-f(\theta)}{2\sin\frac{1}{2}\phi}$. The fact that f is piecewise smooth implies that g is piecewise continuous. Thus,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} g(\phi)\sin(N+\frac{1}{2})\phi\, d\phi \to 0$$

as $N \to \infty$. This is equivalent to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta-\phi)D_N(\phi)\, d\phi \to f(\theta)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(\phi)\, d\phi\right) = f(\theta).$$

It remains to investigate the points where f is discontinuous.

Example 7.1.21. Define

$$g(\theta) = \begin{cases} -1, & \theta \in (-\pi, 0), \\ 0, & \theta = 0, \\ 1, & \theta \in (0, \pi]. \end{cases}$$

Note that $\hat{g}(n) = \frac{1}{in\pi}(1-(-1)^n)$ for $n \neq 0$ and $\hat{g}(0) = 0$. For $\theta \neq 0$, $g(\theta) = \sum_{n\neq 0}\frac{1}{in\pi}(1-(-1)^n)e^{in\theta}$. For $\theta = 0$ observe that the Fourier series sums symmetrically to 0, which is the sum of the left and the right limits of g at 0 divided by 2, that is,

$$\frac{g(0-)+g(0+)}{2}.$$

Exercise 7.1.22. Check the same phenomenon for the Fourier series of $g(\theta) = \theta^N$, for $N \geq 1$ when $\theta = \pi$.

Theorem 7.1.23. *Let f be a piecewise smooth function, let α be a point of discontinuity of f, and let ℓ_+ and ℓ_- denote the right and left limits, respectively, of f at α. Then the Fourier series of f at α converges to $\frac{\ell_++\ell_-}{2}$.*

Proof. Set

$$h(\theta) = f(\theta) - \frac{\ell_+ - \ell_-}{2}g(\theta-\alpha),$$

for $\theta \neq \alpha$, where g is the function in the Example 7.1.21.

Note that

$$\lim_{\theta\to\alpha^-} h(\theta) = \lim_{\theta\to\alpha^+} h(\theta) = \frac{\ell_+ + \ell_-}{2}.$$

Define $h(\alpha)$ to be this common limit. Then h is piecewise smooth on $\mathbb{T}$ and continuous at α. Hence, the Fourier series of h at α sums to $(\ell_+ + \ell_-)/2$. Since the sum of the Fourier series of $g(\theta-\alpha)$ is 0 at α, it follows that the sum of the Fourier series of f at α is the same as that of h.

The next theorem provides important information about the convergence properties of Fourier series.

THEOREM 7.1.24. *Let f be a continuous and piecewise smooth function on $\mathbb{T}$. Then*

(i) *the Fourier series of f converges uniformly on $\mathbb{T}$, and*
(ii) *the Fourier series of f converges absolutely on $\mathbb{T}$.*

Proof. Let $g = f'$ on $\mathbb{T}$. Then g is piecewise continuous and $in\hat{f}(n) = \hat{g}(n)$ for all $n \in \mathbb{Z}$ by Facts 7.1.4. From this we get $\sum_{n=-N}^{N} n^2|\hat{f}(n)|^2 = \sum_{n=-N}^{N} |\hat{g}(n)|^2 \leq \|g\|_2^2$ by Bessel's inequality (Exercise 7.1.6). Thus

$$\sum_{n=-N}^{N} |\hat{f}(n)| = \sum_{n=-N}^{N} \frac{1}{n}(n|\hat{f}(n)|) \leq \left(\sum_{n=-N}^{N} \frac{1}{n^2}\right)^{1/2} \left(\sum_{n=-N}^{N} n^2|\hat{f}(n)|^2\right)^{1/2}$$

by the Cauchy-Schwarz inequality, and absolute convergence follows.

Since f is continuous and piecewise smooth, we know from Theorem 7.1.20 that the sequence $(S_N(f))_{N\in\mathbb{N}}$ converges pointwise to f on $\mathbb{T}$. We will show that this sequence converges uniformly. For $M > N$ we consider

$$\begin{aligned} |S_N(f)(\theta) - S_M(f)(\theta)|^2 &= \left|\sum_{N+1\leq|n|\leq M} \hat{f}(n)e^{in\theta}\right|^2 \\ &\leq \sum_{N+1\leq|n|\leq M} |\hat{f}(n)|^2 \\ &= \sum_{N+1\leq|n|\leq M} \frac{1}{n^2}|\hat{g}(n)|^2 \leq \frac{\|g\|_2^2}{(N+1)^2}. \end{aligned}$$

By choosing M and N large enough, this last term can be made arbitrarily small. This proves the theorem.

EXERCISE 7.1.25. Does either of the above conclusions hold for all piecewise smooth functions on $\mathbb{T}$?

EXERCISE 7.1.26. Compute the Fourier series of $f(\theta) = \theta^n$ on $-\pi < \theta \leq \pi$.

EXERCISE 7.1.27.

(i) Prove that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.
(ii) Prove that $\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}$.
(iii) Prove that $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = \frac{\pi^2}{12}$.
(iv) Prove that $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^3} = \frac{\pi^3}{32}$.

EXERCISE 7.1.28. Suppose that f is piecewise continuous on $\mathbb{T}$. Show that the Fourier series of f can be integrated term by term in the sense that

$$\int_a^b f(\theta)d\theta = \sum_{n=-\infty}^{\infty} \hat{f}(n) \int_a^b e^{in\theta}d\theta.$$

Note that we make no assumptions here about the convergence of the Fourier series itself.

7.1.2. The L^2-Theory.

Suppose that $f \in L^2(\mathbb{T})$. Then $f \in L^1(\mathbb{T})$ and we can consider the Fourier series of f as we did in the previous section. The following exercise sets the tone for Fourier analysis in $L^2(\mathbb{T})$. Recall that the collection χ_n, $n \in \mathbb{Z}$, where $\chi_n(\theta) = e^{in\theta}$, is an orthonormal set in $L^2(\mathbb{T})$. That is,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \chi_n(\theta)\overline{\chi_m(\theta)}\,d\theta = \begin{cases} 0, & n \neq m, \\ 1, & n = m. \end{cases}$$

Note that $S_N(f)$ is in $L^2(\mathbb{T})$ and $\|S_N(f)\|_2^2 = \sum_{n=-N}^{N} |\hat{f}(n)|^2$.

Exercise 7.1.29.

(i) **(Bessel's inequality—once again)** If $f \in L^2(\mathbb{T})$, then $\|f\|_2^2 \geq \|S_N(f)\|_2^2$.

(ii) For a sequence of complex numbers $(a_n)_{n\in\mathbb{Z}}$ and a nonnegative integer N, define $A_N = \sum_{n=-N}^{N} a_n e^{in\theta}$. Show that $\|f - A_N\|_2 \geq \|f - S_N(f)\|_2$.

Sums of the form A_N as in the previous exercise are called *trigonometric polynomials*.

Lemma 7.1.30. *Trigonometric polynomials are dense in $L^2(\mathbb{T})$.*

Proof. Take $f \in L^2(\mathbb{T})$ and $\varepsilon > 0$. As in the case of $L^1(\mathbb{T})$, we choose a function $g \in C^1(\mathbb{T})$ such that $\|f - g\|_2 < \varepsilon/2$. Now choose N so that $\sup|g(\theta) - S_N(g)(\theta)| < \varepsilon/2$. By the triangle inequality, $\|f - S_N(g)\|_2 < \varepsilon$.

Corollary 7.1.31. *Let $f \in L^2(\mathbb{T})$. Then $S_N(f) \longrightarrow f$ in $L^2(\mathbb{T})$.*

Proof. This follows immediately from the previous lemma and Exercise 7.1.29(ii).

Corollary 7.1.32. *The collection $\{\chi_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{T})$.*

Proof. Clear.

Theorem 7.1.33 **(Parseval's identity)**. *If $f \in L^2(\mathbb{T})$, then $\|f\|_2^2 = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2$.*

Proof. This follows from Bessel's inequality and the density of trigonometric polynomials.

Finally we prove the Riesz-Fischer theorem, which provides us with an isometry between $L^2(\mathbb{T})$ and $L^2(\mathbb{Z})$. Here, of course, $L^2(\mathbb{Z}) = \{(a_n)_{n\in\mathbb{Z}} \mid a_n \in \mathbb{C} \text{ and } \sum_{n=-\infty}^{\infty} |a_n|^2 < \infty\}$.

THEOREM 7.1.34 **(Riesz-Fischer)**. *Let $(a_n)_{n\in\mathbb{Z}}$ be a sequence of complex numbers such that $\sum_{n=-\infty}^{\infty}|a_n|^2$ is finite. Then there exists $f\in L^2(\mathbb{T})$ such that $\hat{f}(n)=a_n$ for all $n\in\mathbb{Z}$.*

Proof. Consider the sequence f_N, where $f_N(\theta)=\sum_{n=-N}^{N}a_ne^{in\theta}$. It is clear that $(f_N)_{N\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\mathbb{T})$. Since $L^2(\mathbb{T})$ is complete, there is a function $f\in L^2(\mathbb{T})$ such that $f=\lim_{N\to\infty}f_N$ in L^2. To finish the proof, we must show that $\hat{f}(n)=a_n$ for $n\in\mathbb{Z}$. For this, we observe that, for any $k\in\mathbb{Z}$, $(f|\chi_k)=(f-f_N|\chi_k)+(f_N|\chi_k)$. By the Cauchy-Schwarz inequality, the first term goes to 0 as $N\to\infty$, and the second term is a_k.

EXERCISE 7.1.35. Suppose that f is absolutely continuous (see Definition 2.5.1) and $f'\in L^2(\mathbb{T})$.

(i) Show that $|\hat{f}(0)|\le\|f\|_1$.
(ii) Show that $\sum_{n=-\infty}^{\infty}|n\hat{f}(n)|^2=\|f'\|_2$.

7.2. Fourier Analysis on Locally Compact Abelian Groups

We now provide a general setting for harmonic analysis on locally compact abelian groups, of which Fourier analysis on the circle is a special case. Let G be a locally compact abelian group, and let dx be a (left and right) Haar measure on G. For particular groups, we will want to establish a normalization of dx. For example on $\mathbb{T}$ we normalized $d\theta$ so that $\int_{\mathbb{T}}dx=1/(2\pi)\int_{-\pi}^{\pi}d\theta=1$. The basic harmonics on G are called *characters*. These are the fundamental objects of harmonic analysis.

DEFINITION 7.2.1. A *character* of G is a continuous homomorphism $\chi:G\longrightarrow\mathbb{C}^\times$ where $\mathbb{C}^\times$ is the multiplicative group of the complex numbers. A character of G is called a *unitary character* if $\chi(x)\in\mathbb{T}$ for all $x\in G$. That is, $|\chi(x)|=1$ for all $x\in G$.

There are several easy observations about characters of G.

PROPOSITION 7.2.2. *Let G be a locally compact abelian group.*

(i) *The characters of G form a group under multiplication, called the* character group of G.
(ii) *The unitary characters of G form a subgroup of the character group.*
(iii) *If G is compact, then every character of G is unitary.*

Proof. Suppose that χ_1 and χ_2 are characters of G. Then $(\chi_1\chi_2)(x)=\chi_1(x)\chi_2(x)$. Obviously, $\chi_1\chi_2$ is a character. The character $\chi_0(x)=1$ is the identity, and the inverse of a character χ is $\chi^{-1}(x)=\chi(x)^{-1}$. It is immediate that the product of characters is continuous and the inverse of a character is continuous. Statement (ii) follows easily. Statement (iii) is a consequence of the fact that the continuous image of a compact set is compact and the only compact subgroups of $\mathbb{C}^\times$ are subgroups of $\mathbb{T}$.

EXAMPLES 7.2.3.

(i) Let $G = \mathbb{T}$. The functions $(\chi_n)_{n\in\mathbb{Z}}$ discussed in the last section are characters of $\mathbb{T}$.

(ii) Let $G = (\mathbb{R}, +)$. For any $s \in \mathbb{C}$ define $\chi_s(x) = e^{isx}$. Then χ_s is a character of $(\mathbb{R}, +)$ which is unitary if and only if $s \in \mathbb{R}$.

EXERCISE 7.2.4. Let $G = \mathbb{Q}_p$. For $x \in \mathbb{Q}_p$, $x \neq 0$, write $x = \sum_{k=\nu(x)}^{\infty} a_k p^k$, the p-adic expansion of x. Define

$$\lambda(x) = \begin{cases} \sum_{k=\nu(x)}^{-1} a_k p^k & \text{if } \nu(x) < 0, \\ 0 & \text{if } \nu(x) \geq 0, \end{cases}$$

and take $\lambda(0) = 0$. Now, viewing $\lambda(x)$ as a rational number, set $\chi(x) = e^{2\pi i\lambda(x)}$.

(i) Show that χ is a unitary character of the group $(\mathbb{Q}_p, +)$.

(ii) Take $u \in \mathbb{Q}_p$ and set $\chi_u(x) = \chi(ux)$. Show that χ_u is a unitary character of $(\mathbb{Q}_p, +)$.

Of course, the group $(\mathbb{Q}_p, +)$ is not compact. However, we have the following theorem.

THEOREM 7.2.5. *Every character of $(\mathbb{Q}_p, +)$ is a unitary character.*

Proof. By Exercise 3.4.10(viii), we can write $\mathbb{Q}_p$ as a union of the additive subgroups $\mathfrak{p}^n$, $n \in \mathbb{Z}$. Since $\mathfrak{p}^n$ is compact, any character χ of $(\mathbb{Q}_p, +)$ when restricted to $\mathfrak{p}^n$ must be unitary. Hence χ is unitary.

DEFINITION 7.2.6. Let G be a locally compact abelian group. The *dual group* of G is the group of unitary characters of G. The dual group is denoted by $\hat{G}$.

We now face two major tasks. The first, given a locally compact abelian group G, is to determine its dual group $\hat{G}$. The second task is to topologize $\hat{G}$ in such a way that it becomes a locally compact abelian group. Having accomplished these two tasks, we then turn to harmonic analysis on G. We have already given examples of unitary characters in the cases $G = \mathbb{T}$, $G = (\mathbb{R}, +)$, and $G = (\mathbb{Q}_p, +)$. It is not our intention in this book to give a detailed development of the theory of harmonic analysis on locally compact abelian groups. This theory can be found in several excellent sources, for example, Hewitt and Ross [**9**], [**10**], Katznelson [**11**], Pontryagin [**21**], and Rudin [**26**]. The point is that with the compact open topology (Gel′fand topology), $\hat{G}$ is a locally compact abelian group, and, in the cases that interest us, it is not difficult to determine $\hat{G}$.

There are several simple facts which will be useful for us. For completeness we include the definition of the compact open topology.

DEFINITION 7.2.7. Let G be a locally compact abelian group. The *compact open topology* on $\hat{G}$, the group of unitary characters of G, is defined as

follows. Let K be a compact set in G and let U be an open set in $\mathbb{T}$. The subsets of $\hat{G}$ given by $C(K,U) = \{\chi \in \hat{G} \mid \chi(K) \subseteq U\}$ form a subbasis for the compact open topology on $\hat{G}$.

THEOREM 7.2.8. *Let G be a locally compact abelian group.*

(i) *The group G is compact iff $\hat{G}$ is discrete.*
(ii) *Let G_1, G_2 be locally compact abelian groups. Then $\widehat{G_1 \times G_2}$ is topologically isomorphic to $\hat{G}_1 \times \hat{G}_2$. Recall that topologically isomorphic means isomorphic and homeomorphic.*
(iii) *(Pontryagin-van Kampen Duality Theorem) The natural inclusion $G \longrightarrow \hat{\hat{G}}$ is a topological isomorphism.*

Proof. The proofs of the algebraic aspects of the statements in this theorem are quite simple. It is the topological parts that require a more detailed analysis. These proofs can be found in the references cited above.

REMARK 7.2.9. One of the problems with the general theory of locally compact abelian groups is the absence of a suitable class of "smooth" functions. Such a class of functions, which we designate by V, should have *at least* the following properties.

(1) The class V is a vector space over $\mathbb{C}$.
(2) The space V is dense in $L^p(G)$, $1 \leq p < \infty$.
(3) For $f, g \in V$ we have $f * g \in V$, where $(f * g)(x) = \int_G f(y)g(y^{-1}x)\,dy$ is the convolution.
(4) The space V is closed under translation, that is, if $f \in V$ and $x \in G$, then $f_x \in V$ where $f_x(y) = f(xy)$.

EXAMPLES 7.2.10.

(i) If $G = \mathbb{T}$, we take $V = C^\infty(\mathbb{T})$.
(ii) If $G = (\mathbb{R}, +)$, we take $V = C_c^\infty(\mathbb{R})$, the space of compactly supported infinitely differentiable functions, or $V = \mathcal{S}(\mathbb{R})$, the Schwartz space of infinitely differentiable functions that, along with all of their derivatives, decrease rapidly at infinity. This space will be discussed in Section 7.5.
(iii) If $G = (\mathbb{Q}_p, +)$, we take $V = \mathcal{S}(\mathbb{Q}_p)$, the space of complex-valued locally constant compactly supported functions on $\mathbb{Q}_p$. This space will be discussed in Section 7.6.

The final fact that we deal with is Plancherel's theorem.

DEFINITION 7.2.11. Let G be a locally compact abelian group and let $f \in L^1(G)$. If $\chi \in \hat{G}$, we define the *Fourier transform* of f at χ by $\hat{f}(\chi) = \int_G f(x)\overline{\chi(x)}\,dx$. Thus, the Fourier transform of f is a function on $\hat{G}$.

Since $\hat{G}$ is a locally compact abelian group, we have a Haar measure $d\chi$ on $\hat{G}$.

The two basic theories in harmonic analysis are the smooth theory and the L^2 theory. In the smooth theory, for f in a suitably "nice" class of functions, we can recover f pointwise from its Fourier transform (this is know as *Fourier inversion*). Thus, for suitable f, we have

$$f(x) = \int_{\hat{G}} \hat{f}(\chi)\chi(x)\, d\chi.$$

For the L^2 theory, we wish to establish a relationship similar to Parseval's identity in the case of the circle. Thus, for $f \in L^2(G)$, we wish to prove that $\hat{f} \in L^2(\hat{G})$ and

$$\int_G |f(x)|^2\, dx = \int_{\hat{G}} |\hat{f}(\chi)|^2\, d\chi.$$

This last identity is called Plancherel's theorem. We will have more to say about the origin of that name in the next section.

Observe that in Fourier inversion and Plancherel's theorem, suitable attention must be paid to the normalization of Haar measure. To be specific, in the case when $G = \mathbb{T}$, we have proved Fourier inversion for C^1 functions. That is, if $f \in C^1(\mathbb{T})$, then $f(\theta) = \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{in\theta}$, for all θ. Also, for $f \in L^2(\mathbb{T})$, we have proved that the Fourier series of f converges to f in the L^2 norm. If we normalize Lebesgue measure, $d\theta$, on $\mathbb{T}$ so that $\frac{1}{2\pi}\int_{\mathbb{T}} d\theta = 1$ and if we put the usual counting measure on $\mathbb{Z}$, then $L^2(\mathbb{T})$ is isometric to $L^2(\mathbb{Z})$. So the Plancherel theorem in this case is simply Parseval's identity. That is, $f \in L^2(\mathbb{T})$ if and only if $\hat{f} \in L^2(\mathbb{Z})$ and

$$\frac{1}{2\pi}\int_{\mathbb{T}} |f(\theta)|^2\, d\theta = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2.$$

Before moving on, we make an observation about the Fourier transform $L^2(G)$. If G is not compact, it may be that $L^2(G) \not\subseteq L^1(G)$. In this case, Plancherel's theorem takes the following form.

THEOREM 7.2.12 **(Plancherel's theorem)**. *The Fourier transform on $L^1(G) \bigcap L^2(G)$ is an isometry onto a dense subspace of $L^2(\hat{G})$ and can be extended uniquely to an isometry of $L^2(G)$ onto $L^2(\hat{G})$.*

EXERCISE 7.2.13. Show that if G is an infinite discrete abelian group and $p > q$, then $L^p(G) \subseteq L^q(G)$.

7.3. The Determination of $\hat{G}$

Suppose now that $G = (\mathbb{R}, +)$. For this section we will suppress the plus sign and simply write $G = \mathbb{R}$. We have noted in Example 7.2.3(ii) above that if $t \in \mathbb{R}$, then the function $\chi_t(x) = e^{itx}$ is a unitary character on $\mathbb{R}$. We now show that every unitary character on $\mathbb{R}$ has this form.

EXERCISE 7.3.1. Show that the only closed subgroups H of $(\mathbb{R}, +)$ are $H = \mathbb{R}$, $H = \{0\}$, and $H = \{nb \mid n \in \mathbb{Z}\}$ for some positive $b \in \mathbb{R}$.

THEOREM 7.3.2. *If $G = \mathbb{R}$ and $\chi \in \hat{G}$, then there exists $t \in \mathbb{R}$ such that $\chi = \chi_t$.*

Proof. Let χ be a unitary character on $\mathbb{R}$. Then, if H is the kernel of χ, H is a closed (additive) subgroup of $\mathbb{R}$. By the previous exercise, there are three possibilities: $H = \mathbb{R}$, $H = \{0\}$, or there exists a positive real number b so that $H = \{nb \mid n \in \mathbb{Z}\}$. If $H = \mathbb{R}$, then χ is the identity character, and $t = 0$. The case $H = \{0\}$ is treated in an exercise below. Now, suppose that $H = \{nb \mid n \in \mathbb{Z}\}$ for a positive real number b. We first observe that $\chi(b/2)^2 = \chi(b) = 1$. So $\chi(b/2) = -1$ since $b/2 < b$. Next, $\chi(b/4)^2 = \chi(b/2) = -1$. So, $\chi(b/4) = \pm i$. We may assume that $\chi(b/4) = i$, since otherwise we can simply replace χ by $\overline{\chi}$. We now claim that, for $n > 2$, $\chi(b/2^n) = e^{i\pi/2^{n-1}}$. We already have this for $n = 2$. Assume this is true for some $n \geq 2$ and consider $\chi(b/2^{n+1})$. Notice $\chi(b/2^{n+1})^2 = \chi(b/2^n) = e^{i\pi/2^{n-1}}$. So the choices are $\chi(b/2^{n+1}) = \pm e^{i\pi/2^n}$. If we take the open interval $(-b/4, b/4)$, then its image under χ is the interval $\{e^{i\theta} \mid -\pi/2 < \theta < \pi/2\}$. Consequently, we must have $\chi(b/2^{n+1}) = e^{i\pi/2^n}$. Now take any $x \in [0, b]$ and choose a sequence c_m such that $c_m \to x$ and each term is of the form $c_m = kb/2^n$. Thus $\chi(c_m) = \chi(b/2^n)^k = e^{ik\pi/2^{n-1}}$ and by the continuity of χ this converges to $\chi(x) = e^{i\pi 2x/b}$. Thus, we have $t = 2/b$.

COROLLARY 7.3.3. *If $G = \mathbb{T}$ and $\chi \in \hat{G}$, then there exists $n \in \mathbb{Z}$ such that $\chi = \chi_n$.*

Proof. The group $\mathbb{T}$ is equal to the quotient of $\mathbb{R}$ by the closed subgroup generated by $b = 2\pi$.

EXERCISE 7.3.4. Show that the case $H = \{0\}$ in the above theorem is impossible.

What we have shown here is that if $G = \mathbb{R}$, then $\hat{G} = (\mathbb{R}, +)$ with the operation $\chi_{t_1}\chi_{t_2} = \chi_{t_1+t_2}$. We say that $\mathbb{R}$ is *self-dual*. In addition, if $G = \mathbb{T}$, then $\hat{G} = (\mathbb{Z}, +)$. By Pontryagin duality, $\widehat{(\mathbb{Z}, +)}$ is topologically isomorphic to $\mathbb{T}$, and, for $e^{i\theta} \in \mathbb{T}$, $\chi_\theta(k) = e^{ik\theta}$.

Before we attack the theory of Fourier transforms on the additive group of $\mathbb{R}$, we analyze the dual groups of several important locally compact abelian groups. These facts follow immediately from Theorem 7.3.2 and its corollary.

THEOREM 7.3.5. *Let G be a locally compact abelian group, and let $\hat{G}$ be the character group, or dual group, of G.*

(i) *If $G = (\mathbb{R}_+^\times, \cdot)$, the multiplicative group of the positive real numbers, then $\hat{G}$ is topologically isomorphic to $(\mathbb{R}, +)$. For $t \in \mathbb{R}$, the corresponding character is $\chi_t(x) = x^{it}$, $x \in \mathbb{R}_+^\times$.*

(ii) *If $G = (\mathbb{R}^\times, \cdot)$, the multiplicative group of the nonzero real numbers, then $\hat{G}$ is topologically isomorphic to $(\mathbb{R}, +) \times \mathbb{Z}_2$. For $t \in \mathbb{R}$ and $h \in \{0, 1\}$, we have $\chi_{h,t}(x) = (x/|x|)^h |x|^{it}$.*

(iii) *If $G = (\mathbb{C}, +)$, then $\hat{G}$ is topologically isomorphic to $(\mathbb{C}, +)$. If $w \in \mathbb{C}$, the corresponding character of $(\mathbb{C}, +)$ is $\chi_w(z) = e^{i\mathrm{Re}(z\bar{w})}$.*

(iv) *If $G = (\mathbb{C}^\times, \cdot)$, the multiplicative group of $\mathbb{C}$, then from polar coordinates, we see that G is topologically isomorphic to $\mathbb{R}_+^\times \times \mathbb{T}$. Consequently, $\hat{G}$ is topologically isomorphic to $(\mathbb{R}, +) \times (\mathbb{Z}, +)$. For $t \in \mathbb{R}$ and $n \in \mathbb{Z}$ the corresponding character of $\mathbb{C}^\times$ is $\chi_{n,t}(z) = (z/|z|)^n |z|^{it}$.*

(v) *If $G = \mathbb{T}^n = \mathbb{T} \times \mathbb{T} \times \cdots \times \mathbb{T}$, then $\hat{G}$ is topologically isomorphic to $\mathbb{Z}^n$. For an element $(e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_n}) \in \mathbb{T}^n$, we write $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$. Then, if $m = (m_1, m_2, \ldots, m_n) \in \mathbb{Z}^n$, the corresponding character of $\mathbb{T}^n$ is*

$$\chi_m(\theta_1, \theta_2, \ldots, \theta_n) = e^{im_1\theta_1} \cdot e^{im_2\theta_2} \cdots e^{im_n\theta_n} = e^{i(\mathbf{m}\cdot\theta)}.$$

(vi) *If $G = (\mathbb{R}^n, +)$, then $\hat{G}$ is topologically isomorphic to $(\mathbb{R}^n, +)$. If $t = (t_1, t_2, \ldots, t_n) \in \mathbb{R}^n$ and $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, the corresponding character of $\mathbb{R}^n$ is*

$$\chi_t(x) = e^{i(t_1x_1 + t_2x_2 + \cdots + t_nx_n)} = e^{i(t\cdot x)}.$$

(vii) *If $G = (\mathbb{Q}_p^\times, \cdot)$, then $\mathbb{Q}_p^\times \cong \mathbb{Z} \times U_p$ and U_p is a compact abelian group. Thus $\hat{G} \cong \mathbb{T} \times \hat{U}_p$. Let $\chi \in \hat{G}$ and $x \in \mathbb{Q}_p^\times$ with $x = p^{-\nu(x)}u$, where $u \in U_p$. Then $\chi(x) = |x|^\alpha \psi(u)$, where $\psi \in \hat{U}_p$ and $\alpha = is$ where $s \in \mathbb{R}$ and $-\pi/\ln(p) < s \leq \pi/\ln(p)$. Hence, since $\hat{U}_p$ is a countably infinite group, we see that the unitary dual of $\mathbb{Q}_p^\times$ is a union of a countable number of circles.*

Proof. The proof of (i) follows from the fact that $(\mathbb{R}_+^\times, \cdot)$ is isomorphic to $(\mathbb{R}, +)$ via the map $x \mapsto e^x$. Everything else follows from Theorem 7.3.2, its corollary, and Theorem 7.2.8(ii).

The dual group of $(\mathbb{Q}_p, +)$ is determined in Theorem 7.6.1.

7.4. The Fourier Transform on $(\mathbb{R}, +)$

Now we are ready to consider the Fourier transform on $(\mathbb{R}, +)$.

DEFINITION 7.4.1. For $f \in L^1(\mathbb{R})$ and $t \in \mathbb{R}$, we define the Fourier transform of f at t by

$$\hat{f}(t) = \hat{f}(\chi_t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ixt}\, dx.$$

The constant $1/\sqrt{2\pi}$ appears in front of the integral so that when we consider Fourier inversion on $\mathbb{R}$, the inverse Fourier transform will have the same form. This means we have normalized Haar measure on $(\mathbb{R}, +)$ so that the measure of the closed interval $[0, 1]$ is $1/\sqrt{2\pi}$. With this normalization the Fourier inversion formula states that, for suitable f, we can recover f

pointwise from $\hat{f}$ by

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(t) e^{itx}\, dt.$$

REMARK 7.4.2. Some authors define the Fourier transform by the formula $\hat{f}(t) = \int_{\mathbb{R}} f(x) e^{-itx}\, dx$. In this case, Fourier inversion takes the form $f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t) e^{itx}\, dt$. In fact, rather than use the form of the characters of $\mathbb{R}$ that we derived above, other authors will replace e^{itx} by $e^{2\pi itx}$. This substitution eliminates the need for normalizing constants.

EXERCISE 7.4.3. Let $\mathscr{F}$ denote the Fourier transform operator, that is, $\mathscr{F}(f)(t) = \hat{f}(t)$. Show that $\mathscr{F}^4 = I$.

EXERCISE 7.4.4. Let $f(x) = e^{-x^2/2}$. Show that $\hat{f}(t) = e^{-t^2/2}$. That is, $\mathscr{F}(f) = f$.

As in the case of Fourier series on $\mathbb{T}$, there are simple facts about the Fourier transform on $\mathbb{R}$.

FACTS 7.4.5.

(i) For $f \in L^1(\mathbb{R})$, $\hat{f}$ is bounded on $\mathbb{R}$. In fact, $\|\hat{f}\|_\infty \le \frac{1}{\sqrt{2\pi}} \|f\|_1$.

Proof.

$$|\hat{f}(t)| = \left| \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ixt}\, dx \right| \le \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |f(x)|\, dx.$$

(ii) If $f \in L^1(\mathbb{R})$, then $\hat{f}$ is uniformly continuous.

Proof. For $\delta > 0$,

$$\begin{aligned}
\sqrt{2\pi}|\hat{f}(t+\delta) - \hat{f}(t)| &= \left| \int_{\mathbb{R}} f(x) e^{-itx} (e^{-i\delta x} - 1) dx \right| \\
&\le \int_{\mathbb{R}} |f(x)||e^{-i\delta x} - 1| dx \\
&= \int_{\mathbb{R}} |f(x)| 2|\sin(\delta x/2)| dx \\
&\le 2 \int_{-\infty}^{-R} |f(x)| dx + 2 \int_{R}^{\infty} |f(x)| dx + \delta R \int_{-R}^{R} |f(x)| dx
\end{aligned}$$

since $2|\sin(\delta x/2)| \le \delta R$ for $x \in [-R, R]$. Choose R so that

$$\int_{-\infty}^{-R} |f(x)| dx < \varepsilon/8 \quad \text{and} \quad \int_{R}^{\infty} |f(x)| dx < \varepsilon/8.$$

Then choose $\delta < \varepsilon/(2R\|f\|_1)$ (taking $f \neq 0$).

(iii) If we define $\tilde{f}(x) = \overline{f(-x)}$, then $\widehat{\tilde{f}}(t) = \overline{\hat{f}(t)}$.

(iv) The operator $\mathscr{F}$ is linear; that is, for $f_1, f_2 \in L^1(\mathbb{R})$ and $c_1, c_2 \in \mathbb{C}$ we have

$$\mathscr{F}(c_1 f_1 + c_2 f_2) = c_1 \mathscr{F}(f_1) + c_2 \mathscr{F}(f_2).$$

(v) If $f_1, f_2 \in L^1(\mathbb{R})$, then

$$\int_{\mathbb{R}} f_1(x) \hat{f}_2(x)\, dx = \int_{\mathbb{R}} f_2(x) \hat{f}_1(x)\, dx.$$

Proof. Fubini.

(vi) (Riemann-Lebesgue lemma) If $f \in L^1(\mathbb{R})$, then $\lim_{|t|\to\infty} \hat{f}(t) = 0$.

Proof. First suppose that f is the characteristic function of the interval $[a, b]$. Then it is easy to see that $|\hat{f}(t)| \leq \frac{2}{|t|}$. Then, for any $f \in L^1(\mathbb{R})$, we can approximate f by step functions.

EXERCISE 7.4.6. Suppose that $(f_n)_{n\in\mathbb{N}}$ is a sequence of functions in $L^1(\mathbb{R})$ which converges to f in the L^1 norm. Show that the sequence $(\hat{f}_n)_{n\in\mathbb{N}}$ converges to $\hat{f}$ uniformly on $\mathbb{R}$.

EXERCISE 7.4.7. Let $f \in L^1(\mathbb{R})$.

(i) For $a > 0$, define $f_a(x) = f(ax)$. Show that $\widehat{f_a}(t) = (1/a)\hat{f}(t/a)$.
(ii) For $y \in \mathbb{R}$, define $R_y(f)(x) = f(x + y)$. Show that $\widehat{R_y(f)}(t) = e^{ity}\hat{f}(t)$. For later use, we define $\hat{R}_y = \mathscr{F} R_y \mathscr{F}^{-1}$.
(iii) Show that $\mathscr{F}(L^1(\mathbb{R})) \not\subseteq L^1(\mathbb{R})$. (Hint: Use the characteristic function of $[0, 1]$.)
(iv) Find a function $f : \mathbb{R} \longrightarrow \mathbb{C}$ which is bounded, uniformly continuous, and vanishes at infinity, such that f is not the Fourier transform of a function in $L^1(\mathbb{R})$.

Although the Fourier transform of a function in $L^1(\mathbb{R})$ has a number of nice properties, there is no reasonable way to characterize the space $\{\hat{f} \mid f \in L^1(\mathbb{R})\}$. As we shall see below, for certain spaces of functions, we can completely characterize the Fourier transforms of functions in these spaces.

Throughout analysis, the convolution operator on locally compact groups plays a significant role. Recall that, for $f, g \in L^1(\mathbb{R})$, *the convolution of f and g* is defined by

$$(f * g)(x) = \int_{\mathbb{R}} f(x - y) g(y)\, dy = \int_{\mathbb{R}} f(y) g(x - y)\, dy.$$

THEOREM 7.4.8. *If $f, g \in L^1(\mathbb{R})$, then $f * g \in L^1(\mathbb{R})$ and $\|f * g\|_1 \leq \|f\|_1 \cdot \|g\|_1$. Moreover $\widehat{f * g}(t) = \hat{f}(t)\hat{g}(t)$.*

Proof. This is a simple application of Fubini's theorem.

7.5. Fourier Inversion on $(\mathbb{R}, +)$

We now consider the problem of Fourier inversion, that is, recovering a function f pointwise from $\hat{f}$.

To begin, we take $f \in L^1(\mathbb{R})$, and for $R > 0$ set

$$\begin{aligned} S_R(x) &= \frac{1}{\sqrt{2\pi}} \int_{-R}^{R} e^{itx} \hat{f}(t) dt \\ &= \frac{1}{2\pi} \int_{-R}^{R} \int_{-\infty}^{\infty} f(y) e^{it(x-y)} dy dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x+y) \int_{-R}^{R} e^{-ity} dt dy \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} f(x+y) \frac{\sin(Ry)}{y} dy \\ &= \frac{2}{\pi} \int_{0}^{\infty} \frac{\sin(Ry)}{y} \frac{f(x+y) + f(x-y)}{2} dy. \end{aligned}$$

If we define

$$g_x(y) = \frac{f(x+y) + f(x-y)}{2} - f(x),$$

then

$$S_R(x) - f(x) = \frac{2}{\pi} \int_{0}^{\infty} \frac{\sin(Ry)}{y} g_x(y) dy.$$

EXERCISE 7.5.1. Show that

$$\frac{2}{\pi} \int_{0}^{\infty} \frac{\sin(Ry)}{y} dy = 1.$$

We can now prove our first theorem on Fourier inversion.

THEOREM 7.5.2. *If there exists $\delta > 0$ such that*

$$\int_{0}^{\delta} \left| \frac{g_x(y)}{y} \right| dy < \infty,$$

then

$$\lim_{R \to \infty} S_R(x) = f(x).$$

Proof. Note that

$$S_R(x) - f(x) = \frac{2}{\pi} \left[\int_{0}^{\delta} \sin(Ry) \frac{g_x(y)}{y} dy + \int_{\delta}^{\infty} \sin(Ry) \frac{g_x(y)}{y} dy \right].$$

The second integral goes to 0 as $R \to \infty$ by Corollary 7.1.19. By hypothesis, the first integral goes to 0 as $\delta \to 0$. So, the sum can be made arbitrarily small by first picking δ small enough and then R large enough.

Of course, this result holds only almost everywhere unless we assume continuity or some other smoothness condition on f.

REMARK 7.5.3.

(1) The convergence of $S_R(x)$ to $f(x)$ depends only on the behavior of f in a neighborhood of x. This is called *Riemann's localization principle.*
(2) If f is differentiable at x, then $S_R(x) \to f(x)$ as $R \to \infty$.
(3) If $f \in L^1(\mathbb{R}) \cap C^1(\mathbb{R})$ and $\hat{f} \equiv 0$, then $f \equiv 0$.

EXERCISE 7.5.4. Show that if $f \in L^1(\mathbb{R})$ and $\hat{f} = 0$ a.e., then $f = 0$ a.e.

7.5.1. The Smooth Theory.

Let $f \in C_c^\infty(\mathbb{R})$ with $\text{supp}(f) \subseteq [a, b]$. Then, the Fourier transform of f can be written as a finite integral,

$$\hat{f}(t) = \int_a^b f(x) e^{-itx} dx.$$

EXERCISE 7.5.5. Suppose $f, g \in C_c^\infty(\mathbb{R})$. Show that $f * g \in C_c^\infty(\mathbb{R})$.

DEFINITION 7.5.6. For $f \in C_c(\mathbb{R})$ we define the *Laplace transform* of f at $z \in \mathbb{C}$ by

$$\mathscr{L}(f)(z) = \int_{-\infty}^{\infty} f(x) e^{-izx} dx = \int_a^b f(x) e^{-izx} dx$$

where $\text{supp}(f) \subseteq [a, b]$.

Note that $\mathscr{L}(f)$ is defined for every $z \in \mathbb{C}$ and, in fact, is differentiable at every $z \in \mathbb{C}$. Such functions are referred to as *entire functions* on $\mathbb{C}$.

THEOREM 7.5.7. *If $f \in C_c^\infty(\mathbb{R})$, then $\hat{f}$ cannot be compactly supported on $\mathbb{R}$ unless f (and $\hat{f}$) is identically 0.*

Proof. Suppose $\hat{f}$ is compactly supported on $\mathbb{R}$. Then $\mathscr{L}(f)|_{\mathbb{R}}$ is identically zero outside a finite interval, which implies $\mathscr{L}(f) \equiv 0$, by the identity principle for analytic functions. See [**1**].

As stated in Remark 7.2.9, we want a vector space of smooth functions on $\mathbb{R}$ with several desirable properties. Since $\mathbb{R}$ is self-dual we can also require that this space be invariant under the Fourier transform. This leads us to the Schwartz space.

DEFINITION 7.5.8. The *Schwartz space* on $\mathbb{R}$, $\mathcal{S}(\mathbb{R})$, is the collection of complex-valued functions $f \in C^\infty(\mathbb{R})$ such that f and all its derivatives vanish at infinity faster than any polynomial. That is,

$$\lim_{|x| \to \infty} |x|^k f^{(m)}(x) = 0$$

for all $k, m \in \mathbb{N} \cup \{0\}$.

The second condition states that f and all of its derivatives are *rapidly decreasing* at ∞.

EXAMPLE 7.5.9. The function $f(x) = e^{-x^2} \in \mathcal{S}(\mathbb{R})$.

FACTS 7.5.10.

(1) The space $\mathcal{S}(\mathbb{R})$ is a vector space over $\mathbb{C}$.
(2) The space $C_c^\infty(\mathbb{R})$ (and hence $\mathcal{S}(\mathbb{R})$) is dense in $L^p(\mathbb{R})$, for $1 \leq p < \infty$.
(3) The space $\mathcal{S}(\mathbb{R})$ is not dense in $L^\infty(\mathbb{R})$.
(4) The space $\mathcal{S}(\mathbb{R})$ is an algebra under both pointwise multiplication and convolution.
(5) The spaces $C_c^\infty(\mathbb{R})$ and $\mathcal{S}(\mathbb{R})$ are invariant under translation. That is, if $f \in C_c^\infty(\mathbb{R})$ (or $\mathcal{S}(\mathbb{R})$), then $R_y(f) \in C_c^\infty(\mathbb{R})$ (or $\mathcal{S}(\mathbb{R})$), where $R_y(f)(x) = f(x+y)$.

The proof of the facts above is immediate.

EXERCISE 7.5.11. Show that $C_c^\infty(\mathbb{R})$ and $\mathcal{S}(\mathbb{R})$ are closed under convolution.

EXERCISE 7.5.12. Show that $f \in \mathcal{S}(\mathbb{R})$ iff $\hat{f} \in \mathcal{S}(\mathbb{R})$.

THEOREM 7.5.13. *Fourier inversion holds on $\mathcal{S}(\mathbb{R})$. That is, $f \in \mathcal{S}(\mathbb{R})$ iff $\hat{f} \in \mathcal{S}(\mathbb{R})$, and*

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(t) e^{itx} dt.$$

In particular, $\mathcal{S}(\mathbb{R})$ is invariant under the Fourier transform.

Proof. The theorem follows immediately from Exercise 7.5.12 and Remark 7.5.3(2).

It follows from the above discussion that, for $f \in \mathcal{S}(\mathbb{R})$, we have $\widehat{f'}(t) = it\hat{f}(t)$. In particular, differential operators with constant coefficients go to multiplication by polynomials under the Fourier transform.

7.5.2. Plancherel's Theorem for $\mathbb{R}$.

The Plancherel theorem, as proved by Plancherel, is a theorem about $L^2(\mathbb{R})$. This theorem states that, given $f \in L^2(\mathbb{R})$, one can define a Fourier transform $\hat{f}$ of f in $L^2(\mathbb{R})$ and that $\|f\|_2 = \|\hat{f}\|_2$. In addition, one can recover f from $\hat{f}$ in the L^2 sense. We first illustrate this by working with Schwartz functions.

Let $f \in \mathcal{S}(\mathbb{R})$. As usual, if $\tilde{f}(x) = \overline{f(-x)}$, then $\tilde{f} \in \mathcal{S}(\mathbb{R})$ and $\widehat{\tilde{f}}(t) = \overline{\hat{f}(t)}$. Since $f * \tilde{f} \in \mathcal{S}(\mathbb{R})$, we have

$$\begin{aligned}
(f * \tilde{f})(0) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \widehat{f * \tilde{f}}(t) dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(t) \widehat{\tilde{f}}(t) dt \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |\hat{f}(t)|^2 dt.
\end{aligned}$$

In addition,

$$\begin{aligned}(f * \tilde{f})(0) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)\tilde{f}(-x)dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |f(x)|^2 dx.\end{aligned}$$

For $f \in \mathcal{S}(R)$, we now have $\|f\|_2 = \|\hat{f}\|_2$. Since $\mathcal{S}(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, this result can be extended to $L^2(\mathbb{R})$.

But what does $\hat{f}$ mean for $f \in L^2(\mathbb{R})$? This problem did not arise for $L^2(\mathbb{T})$ since $L^2(\mathbb{T}) \subseteq L^1(\mathbb{T})$ and the Fourier coefficients are well-defined. Nonetheless, we have the following theorem.

THEOREM 7.5.14 **(Plancherel, 1910)**. *Let $f \in L^2(\mathbb{R})$. Define*

$$\hat{f}_N(t) = \frac{1}{\sqrt{2\pi}} \int_{-N}^{N} f(x)e^{-ixt}\, dx.$$

(i) *The sequence $(\hat{f}_N)_{N\in\mathbb{N}}$ is Cauchy in $L^2(\mathbb{R})$. Let $\hat{f}$ be its limit.*
(ii) $\int_{\mathbb{R}} |\hat{f}(t)|^2\, dt = \int_{\mathbb{R}} |f(x)|^2\, dx$.
(iii) *Now define*

$$f_N(x) = \frac{1}{\sqrt{2\pi}} \int_{-N}^{N} \hat{f}(t)e^{ixt}\, dt.$$

Then, the sequence $(f_N)_{N\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\mathbb{R})$ that converges to f in the L^2 norm.

Proof. Exercise.

7.5.3. L^2-Theory.

For $f \in \mathcal{S}(\mathbb{R})$, we have $\|f\|_2 = \|\hat{f}\|_2$, that is,

$$\int_{\mathbb{R}} |f(x)|^2 dx = \int_{\mathbb{R}} |\hat{f}(t)|^2 dt.$$

Here,

$$\hat{f}(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)e^{-itx} dx$$

and

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(t)e^{itx} dt.$$

We also have, for $f, g \in \mathcal{S}(\mathbb{R})$,

$$(f * \tilde{g})(0) = \int_{\mathbb{R}} f(x)\tilde{g}(-x)dx = \int_{\mathbb{R}} f(x)\overline{g(x)}dx = (f|g)$$

and

$$(f * \tilde{g})(0) = \int_{\mathbb{R}} \widehat{f * \tilde{g}}(t)dt = \int_{\mathbb{R}} \hat{f}(t)\overline{\hat{g}(t)}dt = (\hat{f}|\hat{g}).$$

REMARK 7.5.15. If V_1 and V_2 are Hilbert spaces and $T : V_1 \longrightarrow V_2$ is a unitary linear transformation, then T preserves inner products.

Write $\mathscr{F} : \mathscr{S}(\mathbb{R}) \longrightarrow \mathscr{S}(\mathbb{R})$, $\mathscr{F}(f) = \hat{f}$. Then $\mathscr{F}$ is a bijective isometry in the L^2 norm. The Fourier transform $\mathscr{F}$ extends uniquely to a bijective isometry (unitary map) from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$, since $\mathscr{S}(\mathbb{R})$ is dense in $L^2(\mathbb{R})$. If $f \in L^2(\mathbb{R})$, then $\hat{f} \in L^2(\mathbb{R})$ and $\|f\|_2 = \|\hat{f}\|_2$. Moreover, if $f, g \in L^2(\mathbb{R})$, then $(f|g) = (\hat{f}|\hat{g})$. Note that, for $f, g \in L^2(\mathbb{R})$, $\widehat{f * g} = \hat{f}\hat{g}$.

Recall that, in our discussion of Hilbert spaces, we observed that any two infinite-dimensional separable Hilbert spaces are isometrically isomorphic. This was accomplished by mapping an orthonormal basis (ONB) for one space onto an ONB for the second space. In the current context, we have observed that the characters χ_n, $n \in \mathbb{Z}$, form an ONB for $L^2(\mathbb{T})$. This raises the question, "What is an ONB for $L^2(\mathbb{R})$?"

There is a standard ONB for $L^2(\mathbb{R})$ called the *Haar basis*. This is constructed as follows. First, define the Haar mother function

$$\psi(x) = \begin{cases} 1, & 0 \le x < \frac{1}{2}, \\ -1, & \frac{1}{2} \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Now define a bi-indexed series of functions by

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

for $j, k \in \mathbb{Z}$. It is well known (see [**20**]) that the $\psi_{j,k}$ form an ONB for $L^2(\mathbb{R})$.

Exercise 7.5.16. Find an isometry from $L^2(\mathbb{T})$ to $L^2(\mathbb{R})$ that maps the ONB $\{\chi_n\}_{n\in\mathbb{Z}}$ to the Haar basis of $L^2(\mathbb{R})$.

Another technique for finding an ONB for $L^2(\mathbb{R})$ is to first define an isometry from $L^2(\mathbb{T})$ to $L^2(\mathbb{R})$ and use this isometry to determine an ONB for $L^2(\mathbb{R})$. In fact, this isometry emerges from the trigonometric substitutions that occur in elementary calculus. Starting with $f \in L^2(\mathbb{T})$, we define

$$Tf(x) = \frac{1}{\sqrt{\pi}} f\left(\frac{1+ix}{1-ix}\right) |1-ix|^{-1}.$$

We must show that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |f(e^{i\theta})|^2 d\theta = \int_{\mathbb{R}} |Tf(x)|^2 dx = \int_{\mathbb{R}} \left| \frac{1}{\sqrt{\pi}} f\left(\frac{1+ix}{1-ix}\right)\right|^2 |1-ix|^{-2} dx.$$

Using the trigonometric substitution $x = \tan(\theta/2)$ from elementary calculus, the conclusion follows.

The next step is to use the linear operator T to determine an "unusual" ONB for $L^2(\mathbb{R})$. This gives us the following theorem.

Theorem 7.5.17. *The collection*

$$\varphi_n(x) = \frac{1}{\sqrt{\pi}} \left(\frac{1+ix}{1-ix}\right)^n |1-ix|^{-1}$$

for $n \in \mathbb{Z}$ is an ONB for $L^2(\mathbb{R})$.

Proof. This follows from the fact that T is an isometry from $L^2(\mathbb{T})$ to $L^2(\mathbb{R})$.

EXERCISE 7.5.18. Prove directly that $\{\varphi_n\}$ is an ONB for $L^2(\mathbb{R})$. Note that the orthogonality comes easily from trigonometric substitution.

7.6. Fourier Analysis on p-adic Fields

We begin our study of Fourier analysis on p-adic fields with the derivation of the form of the additive characters on $\mathbb{Q}_p$. We then define a space of "smooth" complex-valued functions on $\mathbb{Q}_p$ and discuss Fourier analysis on the additive group. At the end of the section, we discuss Fourier analysis on the multiplicative group $\mathbb{Q}_p^\times$. Many of the facts will be given as exercises because, as stated by one noted expert, "p-adic fields: if you can add, you can integrate."

THEOREM 7.6.1. *Define the* basic additive character *on* $\mathbb{Q}_p$ *by*

$$\psi(x) = \psi_1(x) = e^{2\pi i \lambda(x)}.$$

Then every additive character on $\mathbb{Q}_p$ *has the form* $\psi_u(x)$ *for some* $u \in \mathbb{Q}_p$, *where* $\psi_u(x) = \psi(ux)$. *The dual group* $(\mathbb{Q}_p, +)\hat{}$ *is topologically isomorphic to* $(\mathbb{Q}_p, +)$. *That is,* $(\mathbb{Q}_p, +)$ *is self-dual.*

Proof. We follow the development of Washington [**29**], leaving several parts of the proof as exercises.

EXERCISE 7.6.2. If $z \in \mathbb{T}$ has the property that

$$\lim_{n\to\infty} z^{p^n} = 1,$$

then $z^{p^m} = 1$ for some positive integer m.

Let ψ be a character of $\mathbb{Q}_p$. By continuity we have

$$\lim_{n\to\infty} \psi(1)^{p^n} = \lim_{n\to\infty} \psi(p^n) = \psi(0) = 1,$$

so by the previous exercise, $\psi(1)^{p^m} = 1$ for some positive integer m. Choose an integer b with $0 \le b < p^m$ such that $\psi(1) = e^{2\pi i b/p^m}$. Since $\psi(p^{-1})$ is a p-th root of this number, we have

$$\psi(p^{-1}) = e^{2\pi i p^{-1}(p^{-m}b + b_0)}$$

for some integer b_0 with $0 \le b_0 < p$. Similarly, we have

$$\psi(p^{-2}) = e^{2\pi i p^{-2}(p^{-m}b + b_0 + pb_1)}$$

for some b_1 with $0 \le b_1 < p$. Proceeding in this fashion, we obtain a sequence $b_0, b_1, b_2, \ldots$ of integers between 0 and $p-1$ such that

$$\psi(p^{-k}) = e^{2\pi i p^{-k}(p^{-m}b + b_0 + pb_1 + \cdots + p^{k-1}b_{k-1})}.$$

Define $u \in \mathbb{Q}_p$ by

$$u = p^{-m}b + \sum_{n=0}^{\infty} b_n p^n.$$

EXERCISE 7.6.3.

(i) Show that $\psi(p^{-k}) = \psi_u(p^{-k})$ for any nonnegative integer k.
(ii) Show that $\psi(ap^k) = \psi_u(ap^k)$ for any integers a and k.
(iii) Complete the proof by showing that $\psi = \psi_u$.

EXERCISE 7.6.4. Show that if $u \neq v$, then ψ_u and ψ_v are distinct.

Thus we conclude that $(\mathbb{Q}_p, +)\hat{}$ is equal to $\{\psi_u \mid u \in \mathbb{Q}_p\}$. It remains to establish that the compact open topology on this set is the same as the topology on $\mathbb{Q}_p$. This is left as an exercise for the reader.

In order to do Fourier analysis on $(\mathbb{Q}_p, +)$, we need a measure similar to Lebesgue measure on $\mathbb{R}^n$. This is rather simple. We take a Borel measure dx on the σ-algebra generated by open sets in $\mathbb{Q}_p$. This algebra is generated by sets of the form $a + \mathfrak{p}^n$ where $a \in \mathbb{Q}_p$ and $n \in \mathbb{Z}$, and we normalize the measure by setting

$$\int_{a+\mathfrak{p}^n} dx = p^{-n}.$$

The resulting measure is translation invariant and is characterized among such measures by the condition that the measure of R_p is 1.

LEMMA 7.6.5. *If ψ is the basic additive character, then*

$$\int_{\mathfrak{p}^n} \psi(x)dx = \begin{cases} p^{-n}, & n \geq 0, \\ 0, & n < 0. \end{cases}$$

Proof. If $n \geq 0$, then $\mathfrak{p}^n \subseteq R_p$, so ψ is identically 1 on $\mathfrak{p}^n$, and the result follows from the definition of the measure dx. Suppose that $n < 0$, so that $R_p \subseteq \mathfrak{p}^n$. Choose $a \in \mathfrak{p}^n$ with the property that $\psi(a) \neq 1$ and note that

$$\int_{\mathfrak{p}^n} \psi(x)dx = \int \mathfrak{p}^n \psi(a+x)dx = \psi(a) \int_{\mathfrak{p}^n} \psi(x)dx.$$

The result follows.

COROLLARY 7.6.6. *We have*

$$\int_{|x|=p^n} \psi(x)dx = \begin{cases} (p-1)p^{n-1}, & n \leq 0, \\ -1, & n = 1, \\ 0, & n > 1. \end{cases}$$

Proof. Note that

$$\int_{|x|=p^n} \psi(x)dx = \int_{\mathfrak{p}^{-n}} \psi(x)dx - \int_{\mathfrak{p}^{-n+1}} \psi(x)dx$$

and apply the previous lemma.

As was the case for Fourier analysis on $\mathbb{R}$, we will find it useful to introduce a special space of functions on $\mathbb{Q}_p$ called the *Schwartz space*.

DEFINITION 7.6.7. The *Schwartz space* of $\mathbb{Q}_p$, denoted $C_c^\infty(\mathbb{Q}_p)$, is the space of complex-valued, compactly supported, locally constant functions on $\mathbb{Q}_p$.

The following exercise summarizes some of the basic properties of the Schwartz space.

EXERCISE 7.6.8. Show that the Schwartz space of $\mathbb{Q}_p$ has the following properties.

(i) If $f \in C_c^\infty(\mathbb{Q}_p)$, then f is continuous.
(ii) If $f \in C_c^\infty(\mathbb{Q}_p)$, then f is supported on $\mathfrak{p}^m$ for some m.
(iii) If $f \in C_c^\infty(\mathbb{Q}_p)$, then f is constant on cosets of $\mathfrak{p}^n$ for some n.
(iv) If $f \in C_c^\infty(\mathbb{Q}_p)$, then f takes on only finitely many values (that is, f is a simple function).
(v) The space $C_c^\infty(\mathbb{Q}_p)$ is dense in $L^p(\mathbb{Q}_p)$ for $1 \leq p < \infty$.

We are now ready to define the Fourier transform of an L^1 function on $\mathbb{Q}_p$ and explore its properties.

DEFINITION 7.6.9. For $f \in L^1(\mathbb{Q}_p)$, the *Fourier transform* of f is the function $\hat{f}$ on $\mathbb{Q}_p$ defined by

$$\hat{f}(u) = \int_{\mathbb{Q}_p} f(x)\overline{\psi_u(x)}dx.$$

Note that this definition implicitly exploits the identification of $\mathbb{Q}_p$ with its dual in order to realize $\hat{f}$ is a function on $\mathbb{Q}_p$.

THEOREM 7.6.10. *The map $f \mapsto \hat{f}$ is a bijection of $C_c^\infty(\mathbb{Q}_p)$ onto itself. Moreover, for $f \in C_c^\infty(\mathbb{Q}_p)$, f is supported on $\mathfrak{p}^m$ and constant on cosets of $\mathfrak{p}^n$ if and only if $\hat{f}$ is supported on $\mathfrak{p}^{-n}$ and constant on cosets of $\mathfrak{p}^{-m}$.*

Proof. We verify the second claim. The first will follow from calculations carried out below. Suppose that f is supported on $\mathfrak{p}^m$ and take $y \in \mathfrak{p}^{-m}$. Then

$$\begin{aligned}
\hat{f}(u+y) &= \int_{\mathfrak{p}^m} f(x)\overline{\psi_{u+y}(x)}dx \\
&= \int_{\mathfrak{p}^m} f(x)\overline{\psi(ux+yx)}dx \\
&= \int_{\mathfrak{p}^m} f(x)\overline{\psi(ux)\psi(xy)}dx \\
&= \int_{\mathfrak{p}^m} f(x)\overline{\psi(ux)}dx = \hat{f}(u)
\end{aligned}$$

where we have used the fact that ψ is identically 1 on R_p.

Now assume that f is constant on cosets of $\wp^n$. Then

$$\hat{f}(u) = \int_{\mathbb{Q}_p} f(x)\overline{\psi_u(x)}dx = \int_{\mathbb{Q}_p} f(x+h)\overline{\psi(ux)}dx$$

for any $h \in \mathfrak{p}^n$. Making the change of variables $x \mapsto x - h$, this integral becomes

$$\int_{\mathbb{Q}_p} f(x)\overline{\psi(ux-uh)}dx = \int_{\mathbb{Q}_p} f(x)\overline{\psi(ux)\psi(-uh)}dx = \overline{\psi(-uh)}\hat{f}(u).$$

If $u \notin \mathfrak{p}^{-n}$, then there exists some $h \in \mathfrak{p}^n$ such that $uh \notin R_p$. It follows that $\psi(-uh) \neq 1$ and hence $\hat{f}(u) = 0$.

For any subset $A \subseteq \mathbb{Q}_p$, we denote its characteristic function by χ_A (this is not to be confused with the *character* χ_n of $\mathbb{T}$). But part (iv) of Exercise 7.6.8 implies that any element $f \in C_c^\infty(\mathbb{Q}_p)$ can be written as a finite linear combination of characteristic functions of the form $\chi_{a+\mathfrak{p}^n}$.

The measure dx is constructed to be translation invariant. The following lemma shows how dx transforms under multiplication by a nonzero element of $\mathbb{Q}_p$.

LEMMA 7.6.11. *If $f \in C_c^\infty(\mathbb{Q}_p)$, then*

$$\int_{\mathbb{Q}_p} f(ux)dx = |u|_p^{-1}\int_{\mathbb{Q}_p} f(x)dx.$$

Proof. First observe that $u\mathfrak{p}^n = \mathfrak{p}^{n+\nu(u)}$ and hence

$$\int_{\mathbb{Q}_p} \chi_{\mathfrak{p}^n}(u^{-1}x)dx = \int_{\mathbb{Q}_p} \chi_{u\mathfrak{p}^n}(x)dx = p^{-n-\nu(u)} = p^{-\nu(u)}\int_{\mathbb{Q}_p} \chi_{\mathfrak{p}^n}(x)dx.$$

The result now follows from the translation invariance of dx and the fact that Schwartz functions are simple functions.

In what follows, we denote $\chi_{\mathfrak{p}^n}$, the characteristic function of $\mathfrak{p}^n$, by ξ_n.

LEMMA 7.6.12. *We have $\hat{\xi}_n = p^{-n}\xi_{-n}$.*

Proof. For $u \neq 0$, we have

$$\begin{aligned}\hat{\xi}_n(u) &= \int_{\mathbb{Q}_p} \xi_n(x)\overline{\psi_u(x)}dx = |u|_p^{-1}\int_{\mathbb{Q}_p} \xi_n(u^{-1}x)\overline{\psi(x)}dx \\ &= |u|_p^{-1}\int_{\mathfrak{p}^{n+\nu(u)}} \overline{\psi(x)}dx.\end{aligned}$$

By Lemma 7.6.5, this is 0 or $p^{-n-\nu(u)} = p^{-n}|u|_p$ depending on whether $u \notin \mathfrak{p}^{-n}$ or $u \in \mathfrak{p}^{-n}$, respectively. If $u = 0$, the assertion $\hat{\xi}_n(u) = p^{-n}\xi_{-n}(u)$ is simply the statement that the measure of $\mathfrak{p}^n$ is p^{-n}.

We are now ready to compute the Fourier transform of the characteristic function $\chi_{a+\mathfrak{p}^n}$. Using the change of variables $x \mapsto x + a$, we have

$$\begin{aligned}
\hat{\chi}_{a+\mathfrak{p}^n}(u) &= \int_{\mathbb{Q}_p} \chi_{a+\mathfrak{p}^n}(x)\overline{\psi_u(x)}dx \\
&= \int_{\mathbb{Q}_p} \xi_{\mathfrak{p}^n}(x)\overline{\psi(ux+ua)}dx \\
&= \overline{\psi(ua)}|u|_p^{-1}\int_{\mathbb{Q}_p} \xi_{\mathfrak{p}^n}(u^{-1}x)\overline{\psi(x)}dx \\
&= \overline{\psi(ua)}|u|_p^{-1}\int_{u\mathfrak{p}^n} \overline{\psi(x)}dx \\
&= \overline{\psi(ua)}|u|_p^{-1}\begin{cases} p^{-n-\nu(u)}, & n+\nu(u)\geq 0, \\ 0, & n+\nu(u)<0 \end{cases} \\
&= \begin{cases} \overline{\psi(ua)}p^{-n}, & |u|_p \leq p^n, \\ 0, & |u|_p > p^n. \end{cases}
\end{aligned}$$

Using this result we can express the Fourier transform of any Schwartz function as a linear combination of additive characters on its support.

EXERCISE 7.6.13. Using this decomposition of the Fourier transforms of a Schwartz function, complete the proof of Theorem 7.6.10.

7.7. Independent Projects

7.7.1. The Gibbs Phenomenon.

So far, much of our discussion of Fourier series has focused on how well a Fourier series converges. In this project, our goal is to better understand a major limitation of the convergence of a Fourier series. In particular, we will see that even when a Fourier series converges pointwise to a particular function f, the partial sums S_n may look quite different from f regardless of how large n is.

EXERCISE 7.7.1. Check that the sequence of functions

$$f_n(x) = \begin{cases} nx, & x \in [0, \frac{1}{n}), \\ 2-nx, & x \in [\frac{1}{n}, \frac{2}{n}], \\ 0, & \text{elsewhere} \end{cases}$$

converges to the function $f(x) = 0$ pointwise. In addition, check that, for any n, $f_n(x) = 1$ for some $x \in \mathbb{R}$.

EXERCISE 7.7.2. Find a sequence of continuous rational functions that converges pointwise to 0 such that every function in the sequence takes the value 1 at some point.

A similar phenomenon occurs when we consider certain Fourier series. First, we will examine a series that was studied by Leonhard Euler years before the work of Fourier.

EXERCISE 7.7.3. Show that

$$\sum_{k=1}^{\infty} \frac{\sin(kx)}{k} = \frac{\pi - x}{2}$$

in the interval $(0, 2\pi)$.

The periodic extension of this function is an example of a *sawtooth curve*. As the following exercises will demonstrate, the partial sums of this series differ significantly from the sawtooth curve that they converge to in the vicinity of the discontinuities at values of the form $2\pi k$.

EXERCISE 7.7.4.

(i) Differentiate the partial sum $S_n(x) = \sum_{k=1}^{n} \frac{\sin(kx)}{k}$.
(ii) Using the fact that $2\sin(\frac{1}{2}x)\cos(kx) = \sin(\frac{1}{2}x + kx) + \sin(\frac{1}{2}x - kx)$, rewrite $S_n'(x)$ in a closed form.
(iii) Find the smallest positive value of x such that $S_n'(x) = 0$.
(iv) Show that the value of x that you found corresponds to a maximum value of $S_n(x)$.
(v) From these values, use a computer to evaluate the first maximum value of $S_n(x)$ for $n = 10$, $n = 100$, and $n = 1{,}000$. Note that each of these values is significantly larger than $\frac{\pi}{2}$, the maximum value of $\frac{\pi - x}{2}$. Note further that, for large n, these values exceed $\frac{\pi}{2}$ by about .08949 times the magnitude of the jump discontinuity at 0.

This example illustrates what is known as the Gibbs phenomenon. Generally, the term "Gibbs phenomenon" refers to the fact that the n-th partial sum of the Fourier series for a piecewise smooth function will have large oscillations near a jump discontinuity. Furthermore, the overshoot near a jump discontinuity will not disappear for larger values of n; instead, it approaches a finite limit.

The previous exercises demonstrated that the Gibbs phenomenon exists. The following exercises will allow us to measure the effect of the Gibbs phenomenon more precisely. Note that we are still working with the same function as above.

EXERCISE 7.7.5.

(i) Using the Dirichlet kernel D_n, show that $S_n(x)$ can be approximated by $\int_0^{nx} \frac{\sin t}{t} dt$ for small positive values of x.
(ii) Show that the maximum value of $\int_0^{nx} \frac{\sin t}{t} dt$ occurs at $x = \frac{\pi}{n}$ so that the maximum value is $\int_0^{\pi} \frac{\sin t}{t} dt$.

(iii) Conclude that the ratio of the maximum value of S_n to the maximum value of $f(x) = \frac{\pi - x}{2}$ in the interval $[0, 2\pi]$ approaches

$$\frac{2}{\pi} \int_0^{\pi} \frac{\sin t}{t} dt \approx 1.179$$

as n approaches infinity. In other words, the overshoot due to the Gibbs phenomenon approaches about 8.95% of the size of the jump discontinuity.

Up to this point we have focused our study of the Gibbs phenomenon on a particular sawtooth function. The phenomenon itself, however, is much more general. In fact, it appears whenever there is a jump discontinuity of a piecewise smooth function, and the overshoot always approaches about 8.95% of the size of the jump. The final exercise in this project shows that this reduction to sawtooth functions is enough to prove the more general phenomenon.

EXERCISE 7.7.6.

(i) Let f be a piecewise smooth periodic function with exactly one jump discontinuity in its period. Show that there exist a continuous function g and a sawtooth function h such that $f = g + h$.

(ii) Generalize the previous statement to functions with multiple jump discontinuities in each period.

(iii) Use the previous results to show that the Gibbs phenomenon must occur near every jump discontinuity of a piecewise smooth function.

7.7.2. Induced Representations and the Motion Group.

Suppose that χ is a character of a locally compact abelian group G. We define a representation T on the one-dimensional Hilbert space $V = \mathbb{C}$ by the formula

$$T(g)z = \chi(g)z \quad \text{for } g \in G, z \in \mathbb{C}.$$

EXERCISE 7.7.7. Show that T is a one-dimensional representation of G and that T is unitary if and only if χ is a unitary character of G.

EXERCISE 7.7.8. Using the irreducibility criterion of Theorem 6.8.38, show that every irreducible unitary representation of a locally compact abelian group is one-dimensional.

Thus the irreducible unitary representations of a locally compact abelian group G are identified with the unitary characters of G.

Now let G be a locally compact unimodular group and let B be a closed subgroup of G. We assume that B is abelian. For a unitary character χ of B, define $C_\chi(G)$ to be the space of continuous functions on G that are compactly supported modulo B and satisfy

$$f(bx) = \chi(b)f(x) \quad \text{for all } b \in B, x \in G. \tag{7.1}$$

Here, "compactly supported modulo B" means that there exists a compact subset $C \subseteq G$ such that the support of f is contained in BC.

EXERCISE 7.7.9. Find conditions on G and B under which $C_\chi(G)$ is nonzero for all characters χ. When is $C_\chi(G)$ infinite-dimensional?

For $x \in G$, we define an operator $T_\chi(x)$ on $C_\chi(G)$ by the formula

$$[T_\chi(x)f](y) = f(yx).$$

This action extends uniquely to a unitary representation of G on a Hilbert space V_χ. We illustrate this in a particular case below.

Let

$$G = \left\{ \begin{pmatrix} 1 & z \\ 0 & e^{i\theta} \end{pmatrix} \middle| \, z \in \mathbb{C}, e^{i\theta} \in \mathbb{T} \right\}$$

be the *motion group* of the complex plane. This name comes from the fact that if we represent the complex w by the pair $(1, w)$, we have

$$(1, w) \begin{pmatrix} 1 & z \\ 0 & e^{i\theta} \end{pmatrix} = (1, we^{i\theta} + z),$$

that is, rotation by $e^{i\theta}$ followed by translation by z. Let B and H denote the subgroups

$$B = \left\{ \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix} \middle| \, z \in \mathbb{C} \right\}$$

and

$$H = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix} \middle| \, e^{i\theta} \in \mathbb{T} \right\}$$

of G.

EXERCISE 7.7.10. Show that B is isomorphic to the additive group $(\mathbb{C}, +)$ and H is isomorphic to the group $\mathbb{T}$. Show that G is the semi-direct product of B and H.

EXERCISE 7.7.11. Show that the coset space $B\backslash G$ is compact in the quotient topology.

By Theorem 7.3.5(iii), the characters of B have the form

$$\chi_w \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix} = \chi_w(z) = e^{i\mathrm{Re}(z\overline{w})}$$

for $w \in \mathbb{C}$. For $w \in \mathbb{C}$, let $C_w(G) = C_{\chi_w}(G)$ be the space defined above, and write T_w in place of T_{χ_w}. Note that, by Exercise 7.7.11, the compactness condition can be omitted in the definition of the space $C_w(G)$.

By (7.1), any $f \in C_w(G)$ is determined completely by its restriction to $H \cong \mathbb{T}$. Moreover, any continuous function f on $\mathbb{T}$ can be extended to a function on G defined by

$$f \begin{pmatrix} 1 & z \\ 0 & e^{i\theta} \end{pmatrix} = \chi_w(ze^{-i\theta}) f(e^{i\theta}).$$

EXERCISE 7.7.12. Show that this extension lies in $C_w(G)$ and that this defines a one-to-one correspondence between $C_w(G)$ and $C(\mathbb{T})$.

EXERCISE 7.7.13. Show that, under this one-to-one correspondence, the action of

$$x = \begin{pmatrix} 1 & z \\ 0 & e^{i\theta} \end{pmatrix} \in G$$

on $C_w(G)$ defined above translates to the action of G on $C(\mathbb{T})$ defined by

$$[T_w(x)f](e^{i\varphi}) = \chi_w(ze^{-i(\theta+\varphi)})f(e^{i(\theta+\varphi)}).$$

EXERCISE 7.7.14. Show that the map $T_w(x)$ is an isometry in the L^2 metric and extends to an isometry of the Hilbert space $L^2(\mathbb{T})$. Conclude that the action T_w extends to a unitary representation of G on $L^2(\mathbb{T})$.

EXERCISE 7.7.15.

(i) Show that T_w is irreducible for all $w \in \mathbb{C}$.
(ii) Show that if $|w| = r$, then T_w is unitarily equivalent to T_r.

The irreducible unitary representations T_w are all infinite-dimensional. The remaining irreducible representations of G are all one-dimensional. The latter arise as follows. For any positive integer n, we have the character χ_n of $\mathbb{T}$ defined in Definition 7.1.1. This character defines a one-dimensional representation S_n of G by the formula

$$S_n \begin{pmatrix} 1 & z \\ 0 & e^{i\theta} \end{pmatrix} w = \chi_n(e^{i\theta})w = e^{in\theta}w.$$

EXERCISE 7.7.16. Determine if there are any unitary equivalences among the S_n.

The representations T_w and S_n exhaust the unitary dual of G.

7.7.3. The Weil Representation for $SL_2(\mathbb{R})$.

In this project, we construct a collection of irreducible unitary representations of $SL_2(\mathbb{R})$. As usual, we will first define a collection of Hilbert spaces and then construct a family of operators that constitute the operators of the representations.

Let $\mathcal{S}(\mathbb{C})$ denote the Schwartz space of $\mathbb{C}$. For $n \in \mathbb{Z}$, we consider the subspace $\mathcal{S}_n$ of $\mathcal{S}(\mathbb{C})$ that transforms according to the character $\chi_n(t)$ of $\mathbb{T}$. That is,

$$\mathcal{S}_n = \{f \in \mathcal{S}(\mathbb{C}) \mid f(tz) = \chi_n(t)f(z) \text{ for all } z \in \mathbb{C}, t \in \mathbb{T}\}.$$

Let $V = L^2(\mathbb{C})$ and let V_n denote the closure of $\mathcal{S}_n$ in V.

EXERCISE 7.7.17.

(i) Show that, for each n, V_n is a closed subspace of V.
(ii) Show that if $n, m \in \mathbb{Z}$ and $n \neq m$, then V_n is orthogonal to V_m.

(iii) Show that $V = \bigoplus_{n\in\mathbb{Z}} V_n$.
(iv) Show that each V_n is an infinite-dimensional subspace of V.

We use the V_n as the Hilbert spaces of our family of representations. We first consider the operation of the Fourier transform on these spaces.

EXERCISE 7.7.18. Show that, for each n, the space V_n is invariant under the complex Fourier transform.

Let $\psi(z) = \psi_1(z) = e^{i\mathrm{Re}(z)}$. For

$$\begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \in SL_2(\mathbb{R})$$

define an operator $M(\beta)$ on $\mathcal{S}_n$ by

$$M(\beta)f(z) = \psi(\beta|z|^2)f(z).$$

For

$$\begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix} \in SL_2(\mathbb{R})$$

define an operator $T(\alpha)$ on $\mathcal{S}_n$ by

$$T(\alpha)f(z) = \alpha f(\alpha z).$$

Lastly, define an operator P on $\mathcal{S}_n$ by

$$Pf(z) = -i\hat{f}(z) = \frac{-i}{2\pi}\int_{\mathbb{C}} f(w)\psi_w(z)dw.$$

EXERCISE 7.7.19. Show that the operators $M(\beta)$, $T(\alpha)$, and P do indeed preserve the spaces $\mathcal{S}_n$ for each $n \in \mathbb{Z}$.

Define matrices $U(\beta)$, $S(\alpha)$, and P by

$$U(\beta) = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}, \quad S(\alpha) = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}, \quad \text{and} \quad P = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

EXERCISE 7.7.20. Show that every element of $SL_2(\mathbb{R})$ can be written uniquely in the form $U(\beta)S(\alpha)$ or $U(\beta)S(\alpha)PU(\beta')$. In particular, conclude that $SL_2(\mathbb{R})$ is generated by the matrices defined above.

EXERCISE 7.7.21. Show that the matrices $U(\beta)$, $S(\alpha)$, and P satisfy the following relations:
(i) $P^2 = S(-1)$.
(ii) $U(\beta)U(\beta') = U(\beta + \beta')$.
(iii) $S(\alpha)S(\alpha') = S(\alpha\alpha')$.
(iv) $S(\alpha)U(\beta)S(\alpha^{-1}) = U(\beta\alpha^2)$.

EXERCISE 7.7.22. Show that $SL_2(\mathbb{R})$ is isomorphic to the abstract group generated by the symbols $U(\beta)$ for $\beta \in \mathbb{R}$, $S(\alpha)$ for $\alpha \in \mathbb{R}^\times$, and P, subject to the relations in the previous exercise.

EXERCISE 7.7.23. Show that the operators $M(\beta)$, $T(\alpha)$, and P on $\mathcal{S}_n$ satisfy the relations above, and conclude that they define a representation T_n of $SL_2(\mathbb{R})$ on the vector space $\mathcal{S}_n$.

EXERCISE 7.7.24. Show that T_n extends to a unitary representation (also called T_n) of $SL_2(\mathbb{R})$ on the Hilbert space V_n.

At this point, we must investigate the irreducibility of the representations T_n for $n \in \mathbb{Z}$. Before beginning this, we convert the space $\mathcal{S}_n$ into a space of functions on $\mathbb{R}_+^\times$.

EXERCISE 7.7.25. Show that the map $\mathcal{S}_n \longrightarrow \mathcal{S}(\mathbb{R}_+^\times)$, given by restriction to the positive reals, is an isometric isomorphism of vector spaces. Here, $\mathcal{S}(\mathbb{R}_+^\times)$ is the space of restrictions of functions from $\mathcal{S}(\mathbb{R})$ to $\mathbb{R}_+^\times$.

EXERCISE 7.7.26. Determine a measure $d\mu$ on $R_+^\times$ with respect to which the isomorphism of the previous exercise extends to an *isometric isomorphism*

$$V_n \longrightarrow L^2(\mathbb{R}_+^\times, d\mu)$$

of Hilbert spaces.

EXERCISE 7.7.27. Show that, under this isometric isomorphism, the operators $M(\beta)$ and $T(\alpha)$ correspond to the operators

$$f(x) \longmapsto e^{i\beta x^2} f(x)$$

and

$$f(x) \longmapsto \alpha f(\alpha x),$$

respectively.

If we now take a bounded linear operator on V_n that commutes with the representations T_n, this operator corresponds to a bounded linear operator on $L^2(\mathbb{R}_+^\times)$ that commutes with the operators in the previous exercise. The following three exercises show that the representations T_n on V_n are irreducible when restricted to the subgroup

$$B = \left\{ \left(\begin{array}{cc} \alpha & \beta \\ 0 & \alpha^{-1} \end{array} \right) \middle| \alpha \in \mathbb{R}^\times, \beta \in \mathbb{R} \right\}$$

of upper-triangular matrices in $SL_2(\mathbb{R})$.

EXERCISE 7.7.28. Show that any bounded linear operator on $L^2(\mathbb{R}_+^\times)$ that commutes with multiplication by $e^{i\beta x^2}$ for all $\beta \in \mathbb{R}$ commutes with multiplication by all elements of $L^\infty(\mathbb{R}_+^\times)$.

EXERCISE 7.7.29. Show that any bounded linear operator on $L^2(\mathbb{R}_+^\times)$ that commutes with multiplication by all elements of $L^\infty(\mathbb{R}_+^\times)$ is multiplication by some element of $h \in L^\infty(\mathbb{R}_+^\times)$.

EXERCISE 7.7.30. Show that if multiplication by $h \in L^\infty(\mathbb{R}_+^\times)$ commutes with the operators $T(\alpha)$, then h is constant almost everywhere. Conclude that T_n is irreducible for all $n \in \mathbb{Z}$.

One question must still be answered: "Are there any unitary equivalences in this family of representations?"

EXERCISE 7.7.31. Show that the unitary representations T_n and T_{-n} are unitarily equivalent.

The reader should note that there are other families of irreducible unitary representations of $SL_2(\mathbb{R})$ that are inequivalent to the above representations. In particular, by modifying the additive character ψ, one can obtain another family of irreducible unitary representations that are not equivalent to the representations T_n.

Appendix A

Sets, Functions, and Other Basic Ideas

> Dans la présente Note, on va essayer de préciser une terminologie propre à l'étude des ensembles abstraits. Cette étude a pour but principal d'étendre les propriétés des ensembles linéaires à des ensembles de plus en plus généraux, et par voie de conséquence, de disséquer ces propriétés et d'en rechercher pour chacune la véritable origine. Il en résulte que le fond des démonstrations est pour ainsi dire donné d'avance et que la difficulté consiste en grande partie a préciser dans quel cadre elles viendront se placer. Adopter une terminologie, c'est donc tracer d'avance toute la théorie. Au fur et à mesure du développement de cette théorie, la terminologie a varié et variera encore. Mais il n'est peut-être pas inutile d'en proposer une, adaptée à la situation présente.
>
> – *Maurice René Fréchet,*
> *Extrait des Comptes rendus du Congrès*
> *des Sociétés savantes en 1924.*

In this chapter, we introduce various types of notation to label definitions, theorems, and formulas. We even use examples of sets whose elements are natural numbers before we formally define the natural numbers. We expect the reader will have no problem with this.

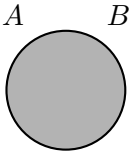
A.1. Sets and Elements

You are probably familiar with the notion of a set as a "collection" or a "bunch" or maybe even a "set" of objects. Formally, we begin our discussion of sets with two undefined terms, that is, "set" and "membership in a set". So we might say that a set is a thing that is a collection of other things called the elements of the set. In practice, this sort of "definition by synonym"

suffices for most mathematicians. If A is a set, we write $x \in A$ to denote membership in A, and we say that x *is an element of the set* A.

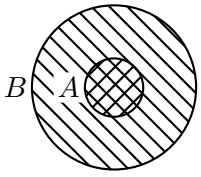
A.2. Equality, Inclusion, and Notation

If A and B are sets, we say that $A = B$ (A *is equal to* B) if they have the same elements. That is, if $x \in A$, then $x \in B$, and conversely if $x \in B$, then $x \in A$. This is used to prove that two sets, which at first glance might not appear to be equal, are indeed equal. Examples of this are given below, and many more are in the exercises. There is one very special set that plays an important role—the *empty set*, which contains no elements. The empty set is denoted $\varnothing$.



$A = B$ $\varnothing$

The next idea is that of a subset. We say that A is a *subset* of B if for any $x \in A$, we have $x \in B$. If A is a subset of B, we write $A \subseteq B$. We also say that B *contains* A (some people even say that B is a *superset* of A, but that is the only time you will see that word in this book). When $A \subseteq B$, it may be the case that B is actually "bigger" than A, that is, there is an element $b \in B$ such that b is not in A, or symbolically, $b \notin A$. In this case, we say that A is a *proper subset* of B, and, if we wish to emphasize this, we write $A \subsetneq B$. However, keep in mind that when we write $A \subseteq B$, A can certainly be a proper subset of B.



$A \subsetneq B$

Exercise A.2.1. If A is a set, show that $A \subseteq A$.

Exercise A.2.2. If A and B are sets, show that $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

Exercise A.2.3. Suppose that A, B, and C are sets. If $A \subseteq B$ and $B \subseteq C$, show that $A \subseteq C$.

Exercise A.2.4. Show that if A is a set, then $\varnothing \subseteq A$.

To be fair, we should observe that all of this is a bit fuzzy logically and may even seem to be tautological. Nonetheless, if you assume the appropriate properties for the symbol $\in$ and if you practice enough, you will feel comfortable with this whole business.

There are two *quantifiers* which we use regularly throughout this book. The first is $\forall$ which reads "for all", and the second is $\exists$ which reads "there exists". Also, the phrase "such that" will often be replaced by the symbol $\ni$, and we abbreviate the phrase "if and only if" by iff.

We usually just assume that all of the sets we consider are contained in some "big" set that is large enough to include all the objects we need. This big set, or *universal set*, is often denoted by the symbol X. Nevertheless, it is possible for a set to be "too big" (see Section A.8). When a quantifier appears without a domain, as in the definition of equality, we mean to consider all objects in the current universe as our domain. Don't get the mistaken idea that the elements of the universal set X must all look the "same". For example, X can contain equilateral polygons, purple Buicks, real numbers, fried green tomatoes, etc.

There is an abbreviated notation for the set of all objects $x \in X$ that satisfy some condition $P(x)$. This notation means that P is a proposition, which is either true or false depending on the value of x. For the set of all $x \in X$ such that $P(x)$ is true, we write $\{x \in X \mid P(x)\}$. We may write simply $\{x \mid P(x)\}$ which again is meant to imply that we take only those x in some designated universe. For example, "the set of x such that x is even" is not sufficiently precise about the universe. It would be better to say, for example, "the set of real numbers x such that x is an even integer".

There will be cases when we list the elements of a set. If the set is small enough, for instance, the set of the first five letters of the alphabet, we write $A = \{a, b, c, d, e\}$. If the set is very large (maybe even infinite), but there is no ambiguity, we may simply list the first few elements of the set and describe the set to the reader. For example, we write the natural numbers as $\mathbb{N} = \{1, 2, 3, 4, \ldots, n, \ldots\}$. This familiar "dot, dot, dot" signifies that you should use your brain and continue as indicated.

A.3. The Algebra of Sets

This section is about taking subsets of a universal set X and putting them together in different ways to create new subsets of X. In fact, that's what most of mathematics is all about, building new things from old things. As was the case in Sections A.1 and A.2, most students will have seen this material, so let's cut right to the chase.

Definition A.3.1. Let A and B be sets. The *union* of A and B, denoted $A \cup B$, is defined by

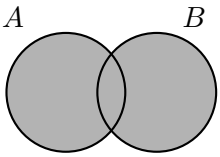
$$A \cup B = \{x \in X \mid x \in A \text{ or } x \in B\}.$$

Note that the "or" in this definition is inclusive (as opposed to "either–or"). That is, even if x is an element of both A and B, then x is still an element of $A \cup B$.

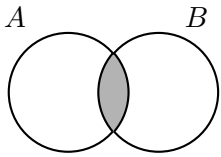
DEFINITION A.3.2. Let A and B be sets. The *intersection* of A and B, denoted $A \cap B$, is defined by

$$A \cap B = \{x \in X \mid x \in A \text{ and } x \in B\}.$$

Here, there is no doubt about what "and" means. "And" is "and".



$A \cup B$

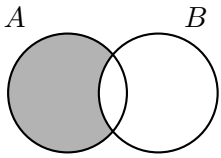


$A \cap B$

DEFINITION A.3.3. Let A and B be sets. We say that A and B are *disjoint* if $A \cap B = \varnothing$. If $\mathcal{C}$ is a collection of sets, any two of which are disjoint, then the elements of $\mathcal{C}$ are said to be *pairwise disjoint*.

DEFINITION A.3.4. Let A and B be sets. The *difference* of A and B, denoted $A \setminus B$ and read "A minus B", is defined by

$$A \setminus B = \{x \in A \mid x \notin B\}.$$



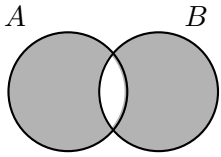
$A \setminus B$

At this point, it is useful to remark that union and intersection are obviously commutative, that is, $A \cup B = B \cup A$ and $A \cap B = B \cap A$. However, difference is not commutative. For example, let $A = \{a\}$ and $B = \varnothing$. The reader may find it amusing to experiment with the difference of various pairs of sets.

The cure for the noncommutativity of the difference is provided by the symmetric difference.

DEFINITION A.3.5. Let A and B be sets. The *symmetric difference* of A and B, denoted $A \triangle B$, is defined by

$$A \triangle B = (A \setminus B) \cup (B \setminus A).$$

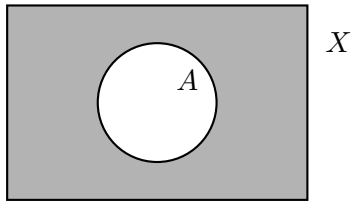


$A \triangle B$

Obviously, the symmetric difference is a commutative operation. Finally, we define the complement of a subset A of a universal set X.

DEFINITION A.3.6. Let $A \subseteq X$. The *complement* of A (in X), denoted ${}^cA$, is defined by

$${}^cA = X \setminus A.$$



${}^cA$

There are many identities among sets that result from using the above operations. We illustrate a few and then assign a multitude of problems for practice.

EXAMPLE A.3.7. This example shows that intersection is distributive over union. That is, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Take $x \in X$. Then $x \in A \cap (B \cup C)$ iff $x \in A$ and $x \in B \cup C$ iff ($x \in A$ and $x \in B$) or ($x \in A$ and $x \in C$). Now this means that $x \in A \cap B$ or $x \in A \cap C$, that is, $x \in (A \cap B) \cup (A \cap C)$. Notice that in this proof, we simply replace symbols by words and use the common understandings of these words.

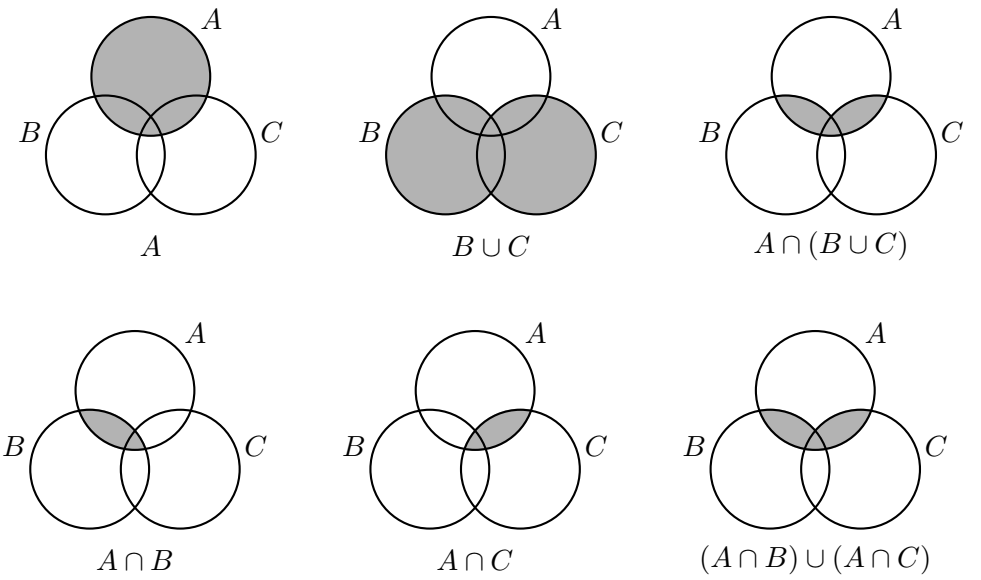


Illustration of the proof of the example

EXAMPLE A.3.8 **(DeMorgan's law. I)**. $^c(A \cup B) = {}^cA \cap {}^cB$.

Take $x \in X$. Then $x \in {}^c(A \cup B)$ iff $x \in X \setminus (A \cup B)$ iff $x \in X$ and $x \notin A \cup B$ iff ($x \in X$ and $x \notin A$) and ($x \in X$ and $x \notin B$) iff $x \in {}^cA$ and $x \in {}^cB$ iff $x \in {}^cA \cap {}^cB$.

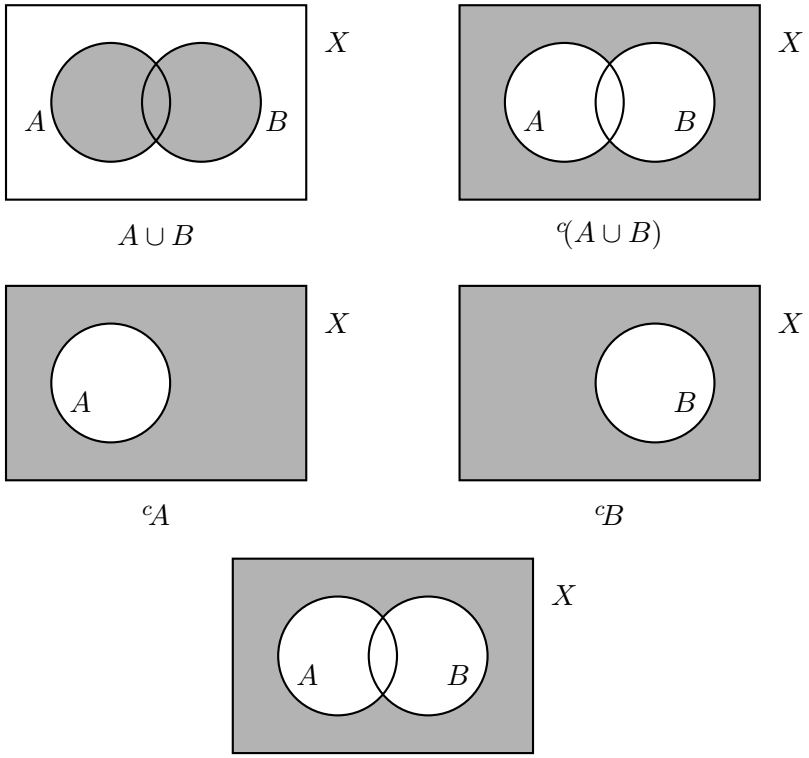


Illustration of the proof of the example.

As the book progresses, occasionally we will need to show, for certain pairs of sets A and B, that $A \subseteq B$ or perhaps that $A = B$. Such inclusions and equalities can be difficult to prove. The following list of exercises will help the reader to develop skills in this direction.

EXERCISE A.3.9. Prove the following equalities. As in Examples A.3.7 and A.3.8, diagrams will illuminate the situation, but they will not suffice for proof. The sets A, B, C are subsets of some universe X.

(i) (Associative law for union) $A \cup (B \cup C) = (A \cup B) \cup C$.
(ii) (Associative law for intersection) $A \cap (B \cap C) = (A \cap B) \cap C$.
(iii) (Associative law for symmetric difference) $A \triangle (B \triangle C) = (A \triangle B) \triangle C$.
(iv) (The empty set is an identity for union) $A \cup \varnothing = A$.
(v) (The empty set is an identity for symmetric difference) $A \triangle \varnothing = A$.
(vi) (The universe is an identity for intersection) $A \cap X = A$.
(vii) $A \cup B = \varnothing$ iff $A = \varnothing$ and $B = \varnothing$.
(viii) $A \cap B = X$ iff $A = X$ and $B = X$.
(ix) $A \triangle B = \varnothing$ iff $A = B$.
(x) (Distributive law of intersection over symmetric difference) $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$.
(xi) (Distributive law of union over intersection) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
(xii) (DeMorgan's law. II). ${}^{c}(A \cap B) = {}^{c}A \cup {}^{c}B$.

It is obvious that the concepts of union and intersection can be extended to any finite number of sets $A_1, A_2, \ldots, A_n$. These are written $\bigcup_{i=1}^{n} A_i$ and $\bigcap_{i=1}^{n} A_i$. More generally, we can extend union and intersection to any collection of sets. We select an *index set* I and write $\bigcup_{i \in I} A_i$ and $\bigcap_{i \in I} A_i$. Here, I could be finite or infinite. This is discussed later in the chapter.

A.4. Cartesian Products, Counting, and Power Sets

The Cartesian product of two sets may be familiar from the example of the coordinate, or Euclidean, plane. This is the set of all pairs (x, y) of real numbers where x denotes the "first coordinate" and y denotes the "second coordinate". The symbol (x, y) is called an *ordered pair*. This is not equal to the ordered pair (y, x), unless $y = x$. That is, the position of the coordinates makes a difference. For example, in the coordinate plane, the point $(1, 2)$ is not the same as the point $(2, 1)$, whereas the sets $\{1, 2\}$ and $\{2, 1\}$ are the same since they have the same elements and order is irrelevant. There is a formal definition of ordered pair, namely $(a, b) = \{\{a\}, \{a, b\}\}$. We are more concerned with a working principle. We say that two ordered pairs (x, y) and (x', y') are equal iff $x = x'$ and $y = y'$.

DEFINITION A.4.1. Let A and B be sets. The *Cartesian product* of A and B, denoted $A \times B$, is defined by

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}.$$

Thus, the Cartesian product of A and B is the set of all ordered pairs in which the first coordinate comes from the set A and the second coordinate comes from the set B. Notice that we have created a new universal set of ordered pairs of elements of the old universal set X. This will cause no difficulty for the moment.

For future use, we make the following definition of a relation on a set X.

DEFINITION A.4.2. A *relation* on a set X is a subset of $X \times X$.

EXAMPLE A.4.3. If $A = \{a, b, c\}$ and $B = \{1, 2, 3\}$, then $A \times B = \{(a, 1), (a, 2), (a, 3), (b, 1), (b, 2), (b, 3), (c, 1), (c, 2), (c, 3)\}$.

EXERCISE A.4.4. Write out $B \times A$ where A and B are as in the example above. Observe that the elements of $B \times A$ are different from those of $A \times B$.

EXERCISE A.4.5. Prove that $A \times \varnothing = \varnothing \times A = \varnothing$.

EXERCISE A.4.6. Suppose $A \neq \varnothing$ and $B \neq \varnothing$. Show $A \times B = B \times A$ iff $A = B$.

There is a fundamental counting principle that accompanies the Cartesian product.

THEOREM A.4.7 (Fundamental counting principle). *If A has m elements and B has n elements, then $A \times B$ has mn elements.*

This is simple to prove by drawing little trees or using some other artifice. A formal proof by induction is very straightforward and will be given as an exercise later in the chapter. This counting principle is the basis for most of the combinatorial formulas in finite probability theory. We will have occasion to use this formula in only a few instances since most of the sets with which we deal in analysis have an infinite number of elements.

We discuss the terms "finite set" and "infinite set" in Section A.8. We denote the number of elements in a set A by $^{\#}A$. So our fundamental counting principle says

$$^{\#}(A \times B) = (^{\#}A)(^{\#}B) \text{ if } A \text{ and } B \text{ are finite.}$$

To generalize this fundamental counting principle to a finite number of finite sets, we must first define the Cartesian product of these sets. Suppose that $A_1, A_2, \ldots, A_n$ are subsets of X.

DEFINITION A.4.8. The n-fold Cartesian product of sets $A_1, \ldots, A_n$ is $A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) \mid a_j \in A_j \text{ for } 1 \leq j \leq n\}$. This is the set of ordered n-tuples, with each coordinate coming from the appropriate subset.

EXERCISE A.4.9. If A_1 has k_1 elements, A_2 has k_2 elements, $\ldots$, A_n has k_n elements, show that $^{\#}(A_1 \times A_2 \times \cdots \times A_n) = (^{\#}A_1)\,(^{\#}A_2) \cdots (^{\#}A_n) = k_1 k_2 \cdots k_n$. Hint: This can be proved drawing pictures but a formal proof is better.

Another counting principle has to do with the union of two sets. When counting the number of elements in $A \cup B$, we cannot simply add $^{\#}A$ and $^{\#}B$ since the intersection might be nonempty (that is, not the empty set), so we would be counting the number of elements in the intersection twice.

EXERCISE A.4.10 (Inclusion-exclusion principle).

(i) If A and B are finite sets and $A \cap B = \varnothing$, show that

$$^{\#}(A \cup B) = {}^{\#}A + {}^{\#}B.$$

(ii) If A and B are finite sets, show that

$$^{\#}(A \cup B) = {}^{\#}A + {}^{\#}B - {}^{\#}(A \cap B).$$

(iii) Do it for three sets; that is, if A, B, and C are finite sets, show that

$$^{\#}(A \cup B \cup C) = {}^{\#}A + {}^{\#}B + {}^{\#}C - {}^{\#}(A \cap B) - {}^{\#}(A \cap C) - {}^{\#}(B \cap C) + {}^{\#}(A \cap B \cap C).$$

(iv) Generalize the previous exercise to any finite number of finite sets.

The next thing to look at is the collection of all subsets of a given set. The idea here is to start with a universe X and study all the subsets of X.

EXERCISE A.4.11.

(i) Let $X = \{1\}$. Write a list of the subsets of X.

(ii) Let $X = \{1, 2\}$. Write a list of the subsets of X.

(iii) On the basis of this information, make a conjecture about the number of subsets of a set with n elements.

DEFINITION A.4.12. Let X be a set. The *power set* of X, denoted $\mathscr{P}(X)$, is the collection of all subsets of X.

Here is the counting principle that goes with $\mathscr{P}(X)$.

THEOREM A.4.13. *If X is a set with n elements, then $\mathscr{P}(X)$ has 2^n elements.*

Proof. Enumerate the elements of X: $x_1, x_2, \ldots, x_n$. Given a subset A of X, we construct a sequence $c_1, c_2, \ldots, c_n$ of 0's and 1's as follows. Let $c_i = 1$ if $x_i \in A$ and $c_i = 0$ if $x_i \notin A$. Thus, the subset A corresponds to a unique sequence of length n consisting of 0's and 1's. Similarly, given a sequence of 0's and 1's of length n, one can construct a unique subset of X. But how many sequences of length n are there consisting of 0's and 1's? By the fundamental counting principle, there are 2^n such sequences.

A.5. Some Sets of Numbers

The set of natural numbers is the collection $\mathbb{N} = \{1, 2, 3, \ldots, n, \ldots\}$. We have already used these numbers to index theorems and examples as well as in examples themselves. Now, we wish to take a more formal approach towards a familiar set of numbers, namely the integers. The integers form the collection $\{0, 1, -1, 2, -2, \ldots\}$ which we study in elementary arithmetic.

We denote the integers by the symbol $\mathbb{Z}$ (from the German word *Zahlen*). The operations in the integers are addition ($+$) and multiplication ($\cdot$), and here are the rules. We expect that the reader is well versed in the arithmetic of the integers, but we are stating these properties explicitly for two reasons. First, these properties are used in arithmetic from the earliest grades but are seldom justified. Second, these properties will be used to describe other algebraic structures that we will meet later.

Rules of Arithmetic in $\mathbb{Z}$ A.5.1.

(A1)	If $a, b \in \mathbb{Z}$, then $a + b \in \mathbb{Z}$.	Closure
(M1)	If $a, b \in \mathbb{Z}$, then $a \cdot b \in \mathbb{Z}$.	Closure
(A2)	If $a, b, c \in \mathbb{Z}$, then $a + (b + c) = (a + b) + c$.	Associativity
(M2)	If $a, b, c \in \mathbb{Z}$, then $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.	Associativity
(A3)	If $a, b \in \mathbb{Z}$, then $a + b = b + a$.	Commutativity
(M3)	If $a, b \in \mathbb{Z}$, then $a \cdot b = b \cdot a$.	Commutativity
(A4)	$\exists 0 \in \mathbb{Z} \ni \forall a \in \mathbb{Z}$, $a + 0 = 0 + a = a$.	Identity
(M4)	$\exists 1 \in \mathbb{Z} \ni 1 \neq 0$ and $\forall a \in \mathbb{Z}$, $a \cdot 1 = 1 \cdot a = a$.	Identity
(A5)	$\forall a \in \mathbb{Z}$, $\exists -a \in \mathbb{Z} \ni a + (-a) = (-a) + a = 0$.	Additive inverses

In general, elements in $\mathbb{Z}$ do not have multiplicative inverses in $\mathbb{Z}$. That is, given an element $a \in \mathbb{Z}$, we cannot necessarily find another element $b \in \mathbb{Z}$ so that $ab = 1$. However, some integers do have multiplicative inverses, namely 1 and -1.

The operations of addition and multiplication are tied together by the distributive law.

(D) If $a, b, c \in \mathbb{Z}$, then $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

Without the distributive law, there would be no connection between addition and multiplication. The richness of the structure is embodied in the interaction between the two operations.

Let's stop and investigate some of the implications of these ten axioms.

Facts A.5.2.

(1) The additive identity is unique.

Proof. Suppose that 0 and $0'$ are additive identities. Then $0 = 0+0' = 0'$.

(2) The multiplicative identity is unique.

Proof. Do it.

(3) Additive inverses are unique.

Proof. Suppose that $a \in \mathbb{Z}$ and $a+a' = 0$. Then $-a+(a+a') = -a+0$. So, by associativity and other things, $((-a)+a)+a' = -a$, and finally $0 + a' = -a$, or $a' = -a$.

(4) (Cancellation for addition) If $a, b, c \in \mathbb{Z}$ and $a + b = a + c$, then $b = c$.

Proof. If $a + b = a + c$, then $-a + (a + b) = -a + (a + c)$, etc.

Question: Why don't we have cancellation for multiplication based on (M1)–(M4)?

Answer: Because we don't have multiplicative inverses (especially for 0).

(5) If $a \in \mathbb{Z}$, then $a \cdot 0 = 0$.

Proof. We can write

$$\begin{aligned} a \cdot 0 &= a \cdot (0 + 0), \\ (a \cdot 0) + 0 &= a \cdot 0 + a \cdot 0 \end{aligned}$$

by properties of the additive identity and the distributive law. Now cancel to get $a \cdot 0 = 0$.

This is really quite something, and it emphasizes the role of the distributive law. What we have here is multiplication by the additive identity reproducing the additive identity. We have more interaction between multiplication and addition in the following statements.

(6) If $a \in \mathbb{Z}$, then $(-1) \cdot a = -a$.

Proof. We can write $a + (-1) \cdot a = 1 \cdot a + (-1) \cdot a = (1 + (-1)) \cdot a = 0 \cdot a = 0$. But additive inverses are unique, so $-a = (-1) \cdot a$.

Notice that, when convenient, we drop the dot which signifies multiplication.

EXERCISE A.5.3. If $a, b \in \mathbb{Z}$, then $(-a)b = a(-b) = -(ab)$.

EXERCISE A.5.4. If $a, b \in \mathbb{Z}$, then $(-a)(-b) = ab$.

Now, what other properties do the integers have? Cancellation for multiplication should be familiar, but in the integers, cancellation for multiplication doesn't follow from the first ten axioms. We now state it as an additional axiom for the integers.

(C) If $a, b, c \in \mathbb{Z}$ with $a \neq 0$ and $ab = ac$, then $b = c$.
Why is $a = 0$ excluded?

EXERCISE A.5.5. Cancellation can be phrased in another way. Show that the statement "if $a, b \in \mathbb{Z}$ and $ab = 0$, then either $a = 0$ or $b = 0$" is equivalent to cancellation.

What else do we have for the integers? We have inequalities. The $<$ sign should be familiar to you. It is a relation on $\mathbb{Z}$ that is subject to the following *rules of order*.

(O1) (Trichotomy) If $a, b \in \mathbb{Z}$, then one and only one of the following holds: $a < b$, $a = b$, or $b < a$.

(O2) (Transitivity) If $a, b, c \in \mathbb{Z}$ with $a < b$ and $b < c$, then $a < c$.

(O3) (Addition) If $a, b, c \in \mathbb{Z}$ and $a < b$, then $a + c < b + c$.
(O4) (Multiplication by positive elements) If $a, b, c \in \mathbb{Z}$, $a < b$, and $0 < c$, then $ac < bc$.

We adopt the usual notation. That is, if $a < b$, we say that "a is less than b". If $a < b$ or $a = b$, we say that "a is less than or equal to b" and write $a \leq b$. If $a < b$, we may also write $b > a$ and say that "b is greater than a". The statement $b \geq a$ is now self-explanatory.

Here are some examples of fundamental exercises and facts which go with the order axioms. For these statements and the following exercises, let $a, b, c \in \mathbb{Z}$.

FACTS A.5.6.

(1) $a > 0$ iff $-a < 0$.

Proof. Suppose $a > 0$. Add $-a$ to both sides.

(2) If $a > 0$ and $b > 0$, then $ab > 0$.

Proof. Suppose $a > 0$. Then, since $b > 0$, $ab > 0 \cdot b = 0$.

(3) If $a > 0$ and $b < 0$, then $ab < 0$.

Proof. Suppose $a > 0$ and $b < 0$. Then $-b > 0$ and $a(-b) = -(ab) > 0$. So $ab < 0$.

(4) If $a < 0$ and $b < 0$, then $ab > 0$.

Proof. If $a < 0$ and $b < 0$, then $-a > 0$ and $-b > 0$. Hence $(-a)(-b) = ab > 0$.

(5) If $a \neq 0$, then $a^2 > 0$.

Proof. If $a > 0$, use fact (2). If $a < 0$, use fact (4).

(6) $1 > 0$.

Proof. $1 = 1^2$.

(7) If $a > b$ and $c < 0$, then $ac < bc$.

Proof. If $a > b$, then $a - b > 0$. Since $-c > 0$, $(-c)(a-b) = -ac + bc > 0$. Hence, $bc > ac$.

(8) If $a > b$, then $-a < -b$.

Proof. Let $c = -1$ and use fact (7).

Are you having fun yet? Good! Try these exercises.

EXERCISE A.5.7. Suppose that $0 < a$ and $0 < b$. Show that $a < b$ iff $a^2 < b^2$.

EXERCISE A.5.8. Suppose that $a < 0$ and $b < 0$. Show that $a < b$ iff $b^2 < a^2$.

EXERCISE A.5.9. Show that $2ab \leq (a^2 + b^2)$.

The set $\mathbb{N}$ of positive elements in $\mathbb{Z}$ is the set of elements that are greater than 0. It is clear that $\mathbb{N}$ is closed under addition and multiplication. If we add trichotomy, these properties lead to an alternate characterization of order.

EXERCISE A.5.10. Suppose now that we have only the first eleven axioms for $\mathbb{Z}$ (that is, through cancellation). Let P be a set of integers with the following properties.

(1) If $a \in \mathbb{Z}$, then one and only one of the following holds: $a \in P$, $a = 0$, or $-a \in P$.
(2) If $a, b \in P$, then $a + b \in P$ and $ab \in P$.

For $a, b \in \mathbb{Z}$, define $a < b$ if $b - a \in P$. Show that this relation satisfies (O1)–(O4). Moreover, if we have a relation that satisfies (O1)–(O4) and if we define $P = \{a \in \mathbb{Z} \mid a > 0\}$, then show that P satisfies properties (1) and (2) above.

EXERCISE A.5.11. Show that the cancellation axiom (C) can be proved using the properties for addition and multiplication and the order axioms.

So far, the integers have five axioms for addition, four for multiplication, one for the distributive law, one for cancellation (which is superfluous), and four for order. There is one more axiom, which plays a crucial role. It is called the *well-ordering principle*. This principle assures us that 1 is the smallest positive integer. This should not come as a surprise, but we do need something to confirm this. In the rational numbers, which we construct later, the first fifteen axioms are satisfied, but there is actually no smallest positive element.

Well-Ordering Principle for $\mathbb{Z}$ A.5.12. If A is a nonempty subset of the positive integers, then A has a least element. That is, there exists an element $a_0 \in A$, such that for all $a \in A$, $a_0 \leq a$.

That does it! We now have fifteen properties that completely characterize the integers. Most of the work with the well-ordering principle will be done later. However, here are a couple of facts which follow immediately from the well-ordering principle.

FACTS A.5.13.

(1) There are no integers between 0 and 1.

Proof. Let $A = \{a \in \mathbb{Z} \mid 0 < a < 1\}$. If $A \neq \varnothing$, then it has a least element a_0 which is in A. So, $0 < a_0 < 1$, and, by property (O4), $0 < a_0^2 < a_0$. But then $a_0^2 \in A$ and a_0 is not the least element.

(2) (Mathematical induction) Let A be a set of positive integers such that $1 \in A$, and if $k \in A$, then $k + 1 \in A$. Then A is the set of all positive integers.

Proof. Suppose there exists a positive integer which is not in A, and let A' be the set of all such positive integers. Then A' is a nonempty subset of the positive integers and hence has a least element c. Now $c > 1$ since $1 \in A$, and there is no integer between 0 and 1. So $c - 1$ is an integer greater than 0. Since $c - 1 < c$, it follows that $c - 1 \in A$, and so $(c - 1) + 1 = c$ is also in A, which is a contradiction.

EXERCISE A.5.14. If n, k are nonnegative integers, we define the *binomial coefficient* $\binom{n}{k}$ by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

where $n! = n(n-1)\cdots 2 \cdot 1$, and we set $0! = 1$. Prove the *binomial theorem*: if $a, b \in \mathbb{Z}$ and n is a positive integer, then

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

(Use mathematical induction.)

REMARK A.5.15. Observe that the binomial coefficient $\binom{n}{k}$ represents the number of ways of choosing k objects from n objects where order does not matter. The binomial coefficient $\binom{n}{k}$ is the number of subsets of k elements in a set with n elements. Of course the binomial theorem implies that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$, the total number of subsets of a set with n elements.

EXERCISE A.5.16.

(i) Prove by induction that if A and B are finite sets, A with n elements and B with m elements, then $A \times B$ has nm elements.
(ii) Prove by induction the corresponding results for k sets.

A.6. Equivalence Relations and the Construction of $\mathbb{Q}$

Recall that a relation on a set X is simply a subset of $X \times X$. For example, we can define a relation on $\mathbb{Z}$ by setting $\mathcal{R}$ equal to $\{(a, b) \mid a, b \in \mathbb{Z} \text{ and } a < b\}$.

Equivalence relations are everywhere in mathematics, and we really mean that. What an equivalence relation does is take a set and partition it into subsets. Some equivalence relations appear to be very natural, some appear to be supernatural, and others appear to make no sense at all.

DEFINITION A.6.1. Let X be a set. An *equivalence relation* on X is a relation $\mathcal{R}$ on X such that the following hold.

(ER1) (Reflexive) For all $a \in X$, $(a, a) \in \mathcal{R}$.
(ER2) (Symmetric) For $a, b \in X$, if $(a, b) \in \mathcal{R}$, then $(b, a) \in \mathcal{R}$.
(ER3) (Transitive) For $a, b, c \in X$, if (a, b) and $(b, c) \in \mathcal{R}$, then $(a, c) \in \mathcal{R}$.

The "twiddle" (or "tilde") notation ($\sim$) is often used in mathematics. That is, if $(a, b) \in \mathcal{R}$, we write $a \sim b$. Then the definition of equivalence relation becomes:

(ER1) (Reflexive) For all $a \in X$, $a \sim a$.
(ER2) (Symmetric) For $a, b \in X$, if $a \sim b$, then $b \sim a$.
(ER3) (Transitive) For $a, b, c \in X$, if $a \sim b$ and $b \sim c$, then $a \sim c$.

Again, speaking loosely, we can refer to $\sim$ as an equivalence relation on X.

EXERCISE A.6.2. Let $\mathcal{R}$ be a relation on X that satisfies the following:
(a) for all $a \in X$, $(a, a) \in \mathcal{R}$,
(b) for $a, b, c \in X$, if $(a, b), (b, c) \in \mathcal{R}$, then $(c, a) \in \mathcal{R}$.
Show that $\mathcal{R}$ is an equivalence relation.

EXAMPLE A.6.3. The most basic example of an equivalence relation is equality. That is, $a \sim b$ iff $a = b$. Prove this, but please don't write anything.

EXAMPLE A.6.4. If A and B are triangles in the plane, write $A \sim B$ if and only if A is similar to B.

EXAMPLE A.6.5. Let n be an integer greater than or equal to 2. If $a, b \in \mathbb{Z}$, we say that $a \sim b$ iff $a - b$ is a multiple of n, that is, n divides $a - b$.

This last example requires a little more elucidation. So, we present a brief discussion about divisibility in $\mathbb{Z}$.

DEFINITION A.6.6. Suppose that a and b are integers. We say that a *divides* b, written $a|b$, if there is an element $c \in \mathbb{Z}$ such that $b = ac$. The number a is called a *divisor* of b.

We need the following facts about divisibility.

FACTS A.6.7.
(1) If $a \in \mathbb{Z}$, then $a|a$.
(2) If $a|b$, then $a| - b$.
(3) If $a|b$ and $b|c$, then $a|c$.

These facts are easy to prove. For example, if $a|b$ and $b|c$, there are integers h and k such that $b = ha$ and $c = kb$. But then $c = (hk)a$, and that does it.

EXERCISE A.6.8. Show that if $a \in \mathbb{Z}$, then $a|0$.

EXERCISE A.6.9. Show that if a and b are integers such that $a|b$ and $b|a$, then $a = \pm b$.

EXERCISE A.6.10. Show that if $c|a$ and $c|b$ and $s, t \in \mathbb{Z}$, then $c|(sa + tb)$.

There is one other type of integer which should be familiar to the reader.

DEFINITION A.6.11. Let p be a positive integer greater than or equal to 2. We say that p is *prime* if the only positive divisors of p are 1 and p.

If n is a positive integer greater than 2 that is not prime, then n is called *composite*. So, if n is composite, there exist integers a and b both greater than or equal to 2, such that $n = ab$.

EXERCISE A.6.12. Let n be a positive integer greater than or equal to 2. Then there exists a prime p such that p divides n.

The partitioning into subsets relative to an equivalence relation comes about as follows. If $a \in X$, we write $C(a) = \{b \in X \mid b \sim a\}$. The set $C(a)$ is called *the class of* a or *the equivalence class containing* a. Here are the properties of equivalence classes.

THEOREM A.6.13 (Properties of equivalence classes).

(1) $a \in C(a)$.
(2) *If* $a \sim b$, *then* $C(a) = C(b)$.
(3) *If* a *is not equivalent to* b $(a \nsim b)$, *then* $C(a) \cap C(b) = \varnothing$.
(4) $\bigcup_{a \in X} C(a) = X$.

Proof. (1) Reflexivity.
(2) Transitivity.
(3) If $c \in C(a) \cap C(b)$, then $c \sim a$ and $c \sim b$, so $a \sim b$. So we conclude that $C(a) \cap C(b) \neq \varnothing$ iff $C(a) = C(b)$.
(4) Use (1) above.

This all means that an equivalence relation on a set X partitions X into a collection of pairwise disjoint subsets. Although this looks quite special, it's really not that impressive. For example, take a set X and break it up into pairwise disjoint nonempty subsets whose union is all of X. Then, for $a, b \in X$, define $a \sim b$ if a and b are in the same subset.

EXERCISE A.6.14. Prove that this is an equivalence relation on X.

One more important example of an equivalence relation gives a method for constructing the rational numbers $\mathbb{Q}$ from the integers $\mathbb{Z}$ using the properties discussed in the last section. We consider the set $F = \{(a, b) \mid a, b \in \mathbb{Z} \text{ and } b \neq 0\}$. We are thinking (for example) of the pair $(2, 3)$ as the fraction $2/3$. For $(a, b), (c, d) \in F$, we define $(a, b) \sim (c, d)$ if $ad = bc$. Thus, for instance, $(2, 3) \sim (8, 12) \sim (-6, -9)$.

EXERCISE A.6.15. Show that $\sim$ is an equivalence relation on F.

The set of equivalence classes determined by this equivalence relation is called the *rational numbers* and is denoted by $\mathbb{Q}$. You should be extremely happy about this since it explains all that business about equivalent fractions that you encountered in elementary school. What a relief!

We have several things to do with this example. First, we have to add and multiply rational numbers, that is, add and multiply equivalence classes. The fundamental principle to be established here is that, when we add or multiply equivalence classes, we do it by selecting an element from each equivalence class and adding or multiplying these. We must be certain that the result is independent of the representatives that we choose in the

equivalence classes. For simplicity, we denote the class of (a, b) by $\{(a, b)\}$ rather than $C((a, b))$.

For $\{(a, b)\}, \{(c, d)\} \in \mathbb{Q}$, we define $\{(a, b)\} + \{(c, d)\} = \{(ad + bc, bd)\}$ and $\{(a, b)\} \cdot \{(c, d)\} = \{(ac, bd)\}$. What we must establish is the fact that if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$, then $(ad + bc, bd) \sim (a'd' + b'c', b'd')$ and $(ac, bd) \sim (a'c', b'd')$. All this requires is a little elementary algebra, but, for your sake, we'll actually do one and you can do the other. Of course, we do the easier of the two and leave the more complicated one for you. So, here goes: $(a, b) \sim (a', b')$ means that $ab' = a'b$, and $(c, d) \sim (c', d')$ means that $cd' = c'd$. Multiplying the first equality by cd' and then substituting $cd' = c'd$ on the right-hand side of the resulting equation, we get the desired equality $acb'd' = a'c'bd$.

EXERCISE A.6.16. You do addition. It's messy.

When we are defining some operation which combines equivalence classes, we often do this by choosing representatives from each class and then showing that it doesn't make any difference which representatives are chosen. We have a formal name for this. We say that the operation under consideration is *well-defined* if the result is independent of the representatives chosen in the equivalence classes.

Throughout this book, we will encounter equivalence relations on a regular basis. You will be fortunate enough to have the opportunity to prove that these are actually equivalence relations.

What properties are satisfied by addition and multiplication as defined above? For example, what about the associativity of addition? We must prove that $(\{(a, b)\} + \{(c, d)\}) + \{(e, f)\} = \{(a, b)\} + (\{(c, d)\} + \{(e, f)\})$. Well,

$$\begin{aligned}(\{(a, b)\} + \{(c, d)\}) + \{(e, f)\} &= \{(ad + bc, bd)\} + \{(e, f)\} \\ &= \{((ad + bc)f + (bd)e, (bd)f)\}.\end{aligned}$$

Now we use associativity and distributivity in $\mathbb{Z}$ to rearrange things in an appropriate fashion. This gives $\{(((ad)f + (bc)f) + (bd)e, (bd)f)\}$, and using the acrobatics of parentheses, we get $\{(a(df) + b(cf + de), b(df))\} = \{(a, b)\} + (\{(c, d)\} + \{(e, f)\})$. This is all rather simple. To prove various properties of addition and multiplication in $\mathbb{Q}$, we reduce them to known properties from $\mathbb{Z}$.

EXERCISE A.6.17.

(i) Prove the associative law for multiplication in $\mathbb{Q}$.
(ii) Prove the commutative laws for addition and multiplication in $\mathbb{Q}$.
(iii) Show that $\{(0, 1)\}$ is an additive identity in $\mathbb{Q}$.
(iv) Show that $\{(1, 1)\}$ is a multiplicative identity in $\mathbb{Q}$.
(v) Show that $\{(-a, b)\}$ is an additive inverse for $\{(a, b)\}$.
(vi) Prove the distributive law for $\mathbb{Q}$.

Notice here that if $\{(a,b)\} \neq \{(0,1)\}$, that is, $a \neq 0$, then $\{(a,b)\} \cdot \{(b,a)\} = \{(1,1)\}$. Thus, in $\mathbb{Q}$, we have multiplicative inverses for nonzero elements.

Let's tidy this up a bit. First of all, we have no intention of going around writing rational numbers as equivalence classes of ordered pairs of integers. So let's decide once and for all to write the rational number $\{(a,b)\}$ as a/b. Most of the time this fraction will be reduced to lowest terms, but if it is not reduced to lowest terms, it will certainly be in the same equivalence class as a fraction which is reduced to lowest terms. With this, addition and multiplication of rational numbers have their usual definition:

$$\begin{aligned} \frac{a}{b} + \frac{c}{d} &= \frac{ad+bc}{bd}, \\ \frac{a}{b} \cdot \frac{c}{d} &= \frac{ac}{bd}. \end{aligned}$$

Now consider the axioms for the integers (A1)–(A5), (M1)–(M4), and (D). All of these hold for the rational numbers, and there is another multiplicative property, multiplicative inverses.

(M5) If $a \neq 0$, then there is an element a^{-1} such that $aa^{-1} = a^{-1}a = 1$.

The operations of addition and multiplication are sometimes called *binary operations*, or *internal laws of composition*.

Definition A.6.18. Let R be a nonempty set. An *internal law of composition (ILC)* on R is a map $\circ : R \times R \longrightarrow R$. If $a, b \in R$, then we usually write $\circ((a,b)) = a \circ b$.

Of course, the more properties that are satisfied by internal laws of composition, the better life gets.

Definition A.6.19. A set with two internal laws of composition, $+$ and $\cdot$, that satisfy (A1)–(A5), (M1)–(M4), and (D) is called a *commutative ring with* 1. If, in addition, cancellation (C) holds for multiplication, the commutative ring with 1 is called an *integral domain*. If (M5) also holds, the structure is called a *field*.

Note that the word "commutative" in this definition refers not to the commutativity of addition but to the commutativity of multiplication. Thus, in our latest terminology, $\mathbb{Z}$ is an integral domain and $\mathbb{Q}$ is a field. What about cancellation for multiplication? This followed from order in $\mathbb{Z}$, but for $\mathbb{Q}$ (or any field for that matter) cancellation for multiplication holds automatically.

Exercise A.6.20. Prove this.

Exercise A.6.21. Let X be a nonempty set and let $R = \mathscr{P}(X)$. Show that R with symmetric difference as addition and intersection as multiplication is a commutative ring with 1. When is R a field?

There is another definition which will prove useful in our discussions about these various algebraic structures.

DEFINITION A.6.22. Suppose that R is a commutative ring with 1. A subset R_0 of R is a *subring* if R_0 is a ring itself with the same operations of addition and multiplication as in R. We don't necessarily require that R_0 have a multiplicative identity. In the case when it does not, we call R_0 simply a *commutative ring*.

The same idea can be used to define *subintegral domain*. Finally, if F is a field and F_0 is a subset of F, we say that F_0 is a *subfield* if it is a field with the same operations of addition and multiplication as in F.

EXERCISE A.6.23.

(i) Let R be a ring and let R_0 be a nonempty subset of R. Show that R_0 is a subring iff, for any $a, b \in R_0$, we have $a - b$ and ab in R_0.
(ii) If F is a field and F_0 is nonempty subset of F, are the properties in (i) enough to ensure that F_0 is a subfield?

What about order in $\mathbb{Q}$? It is simple to extend the order from $\mathbb{Z}$ to $\mathbb{Q}$. We do this using the notion of a set of positive elements. We say that $a/b \in \mathbb{Q}$ is positive if $ab > 0$ in $\mathbb{Z}$.

EXERCISE A.6.24. Show that the above notion of positivity in $\mathbb{Q}$ satisfies the properties in Exercise A.5.10, or equivalently, the properties of order given in (O1)–(O4).

DEFINITION A.6.25. An integral domain or field in which there is an order relation satisfying (O1)–(O4) is called an *ordered integral domain* or *ordered field*, respectively. See the project in Section A.10.3 for more about this.

REMARK A.6.26. Note that the natural numbers $\mathbb{N}$ may be regarded as a subset of $\mathbb{Z}$, and in turn the integers $\mathbb{Z}$ may be regarded as a subset of $\mathbb{Q}$ by identifying the integer n with the equivalence class $n/1$.

So what is this all about? We have rules for the integers, and the same rules, along with (M5), are satisfied by the rational numbers. Actually, there are lots of structures other than the integers and the rational numbers that have operations of addition and multiplication, and some of them also carry an order relation.

We want to give two more examples before we leave this section. First, let n be a positive integer greater than or equal to 2 and consider the equivalence relation given in Example A.6.5. What are the equivalence classes?

For example, take $n = 5$. Then we have five classes. They are

$$\begin{aligned} C(0) = \overline{0} &= \{0, 5, -5, 10, -10, \dots\}, \\ C(1) = \overline{1} &= \{1, 6, -4, 11, -9, \dots\}, \\ C(2) = \overline{2} &= \{2, 7, -3, 12, -8, \dots\}, \\ C(3) = \overline{3} &= \{3, 8, -2, 13, -7, \dots\}, \\ C(4) = \overline{4} &= \{4, 9, -1, 14, -6, \dots\}. \end{aligned}$$

Note that, in this example, we have simplified the notation of equivalence class by denoting the equivalence class $C(a)$ by $\bar{a}$. Observe that $\overline{5} = \overline{0}$, $\overline{6} = \overline{1}$, etc. In general, for an arbitrary n, we will have n classes $\overline{0}, \overline{1}, \dots, \overline{n-1}$. These are called *the equivalence classes modulo* n, or, for short, *mod* n. Moreover, for any integer a, we denote the equivalence class in which a lies by $\overline{a}$. Of course, it is always true that $\overline{a}$ is equal to one of the classes $\overline{0}, \overline{1}, \dots, \overline{n-1}$. Let's define addition and multiplication mod n.

DEFINITION A.6.27. Denote the set of equivalence classes $\overline{0}, \overline{1}, \dots, \overline{n-1}$ by $\mathbb{Z}_n$. For $\overline{a}, \overline{b} \in \mathbb{Z}_n$, define $\overline{a} + \overline{b} = \overline{a+b}$ and $\overline{a}\,\overline{b} = \overline{ab}$.

EXERCISE A.6.28.

(i) Show that addition and multiplication in $\mathbb{Z}_n$ are well-defined.
(ii) Show that, with these operations, $\mathbb{Z}_n$ is a commutative ring with 1.
(iii) Show that $\mathbb{Z}_n$ cannot satisfy the order axioms no matter how $>$ is defined.
(iv) Show that $\mathbb{Z}_2$ is a field but $\mathbb{Z}_4$ is not.
(v) For p prime, show that $\mathbb{Z}_p$ is a field.

The second example is the real numbers denoted by $\mathbb{R}$. A construction and complete discussion of the real numbers is given in Chapter 1. For the moment, however, it will suffice to say that the real numbers are an ordered field that contains $\mathbb{Q}$ and has one additional property called the least upper bound property. In Appendix B, we use the real numbers as an example without being concerned with this additional property.

A.7. Functions

If you think equivalence relations are everywhere, wait until you see functions. We would all be better off if functions were introduced in kindergarten and studied regularly thereafter. The concept of a function is one of the most important ideas in mathematics. We give the informal definition first because it is much closer to the way people think about functions in practice.

Informally, a function from a set A to a set B is a correspondence between elements of A and elements of B such that each element of A is associated to exactly one element of B. This includes the familiar numerical functions of calculus, where, most often, the sets A and B are the real numbers or

subsets thereof. But it also includes many examples which have nothing to do with the concept of numbers.

EXAMPLE A.7.1. Given any set A, there is a unique function from A to A that assigns each element of A to itself. This is called the *identity function* on A.

What functions do is take elements of a given set and push them into another set (or maybe even the same set). The requirement is that to each element of the first set there must correspond exactly one element of the second. This does not preclude having two distinct elements of the first set correspond to the same element of the second set.

EXAMPLE A.7.2. Let A and B be nonempty sets and choose a fixed element $b \in B$. Define a function from A to B by letting every element of A correspond to b. This is called a *constant function*.

Before we go too far with the informal idea, let's give a more formal definition for the notion of function.

DEFINITION A.7.3. Let A and B be nonempty sets. A *function* from A to B is a subset of $A \times B$ such that each element of A occurs exactly once as a first coordinate.

This, of course, is an entirely useless definition, but it does carry with it the idea expressed informally above. That is, to each element of A there corresponds exactly one element of B. When you think of functions, you will hardly ever think of ordered pairs. The informal notion of a correspondence satisfying certain properties should be your guide. Here's the notation we use. If A and B are sets and f is a function from A to B, we write $f : A \longrightarrow B$. If $a \in A$, we write $f(a)$ for the corresponding element of B. So, if we were to write this as an ordered pair, we would write $(a, f(a))$. We also use the notation $a \mapsto f(a)$, read as "a goes to $f(a)$", to indicate that $f(a)$ corresponds to a.

EXERCISE A.7.4. How would you formulate the definition of function if either A or B were the empty set?

EXAMPLE A.7.5. Take $A = \{a, b, c, d, e\}$ and $B = \{1, 2, 3, 4\}$. Now consider functions from A to B; that is, assign a number to each letter. For example one such function is $\{(a, 1), (b, 2), (c, 3), (d, 4), (e, 2)\}$.

EXERCISE A.7.6. Determine all the functions from A to B in the previous example.

EXERCISE A.7.7.

(i) If A has n elements and B has m elements, how many functions are there from A to B?

(ii) Let $B = \{0, 1\}$. Use the conclusion of part (i) to give an alternate proof of Theorem A.4.13.

This is a convenient place to state the *pigeonhole principle*.

THEOREM A.7.8 (Pigeonhole principle). *Suppose that m and n are positive integers with $n > m$. If n objects are distributed in m boxes, then some box must contain at least two objects. In terms of functions, the pigeonhole principle can be stated as follows. Suppose that A is a set with n elements, B is a set with m elements, and $n > m$. If $f : A \longrightarrow B$ is a function, there are at least two distinct elements of A that correspond to the same element of B.*

EXERCISE A.7.9. Prove this any way you choose.

We turn next to the language of functions. Here is a list of important terms. Let f be a function from A to B. That is, $f : A \longrightarrow B$.

DEFINITIONS A.7.10.

(a) The set A is called the *domain* of f.
(b) If $A' \subseteq A$, we define $f(A') = \{b \in B \mid \exists a \in A' \text{ with } f(a) = b\}$. The set $f(A')$ is called the *image of A' in B under f*. In particular, $f(A)$ is the *image of A in B under f*. This is commonly called the *image* of f.
(c) Note that there is no reason in the world for thinking that $f(A) = B$. If $f(A) = B$, we say that f is *onto* or *surjective* (f is a *surjection*). In general, B is called the *range* of f. Note that if we change the range to $f(A)$, then f is surjective. That is, a function is always surjective onto its image.
(d) Along with the property of being a function, f may also have the property that each element in the image of f corresponds to exactly one element in A. This can be written as follows. For any $a, a' \in A$ if $f(a) = f(a')$, then $a = a'$. A function with this property is called *one-to-one* or *injective* (f is an *injection*).
(e) A function that is one-to-one and onto (that is, injective and surjective) is called *bijective* (f is a *bijection*). A bijection between two sets is often called a *one-to-one correspondence* between the sets.
(f) Two functions f and g are the same, or equal, when they have the same domain and same range and the sets of pairs $\{(a, f(a))\}$ and $\{(a, g(a))\}$ are identical.

EXERCISE A.7.11. Determine which of your functions in Exercise A.7.6 are surjective. How many are there? Notice that there are no injections or bijections. Why is this?

There is a way to combine functions which is very useful for many purposes.

DEFINITION A.7.12. Suppose A, B, and C are sets and $f : A \longrightarrow B$ and $g : B \longrightarrow C$ are functions. The *composition* of f and g is the function $g \circ f : A \longrightarrow C$ defined by $(g \circ f)(a) = g(f(a))$.

Of course, you met composition of functions in elementary calculus, and you enjoyed learning the chain rule. There is one very useful property concerning composition of functions, that is, composition of functions is *associative*.

THEOREM A.7.13. *If A , B , C, and D are sets and if $f : A \longrightarrow B$, $g : B \longrightarrow C$, and $h : C \longrightarrow D$ are functions, then $h \circ (g \circ f) = (h \circ g) \circ f$.*

Proof. Suppose $a \in A$. Then $(h \circ (g \circ f))(a) = h((g \circ f)(a)) = h(g(f(a))) = (h \circ g)(f(a)) = ((h \circ g) \circ f)(a)$.

Let's stop for a minute with the definitions and consider some numerical examples. The sets we work with will be the natural numbers $\mathbb{N}$, the integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$. (You may know what the real and complex numbers are, but we will construct them in Chapter 1.) Often, when writing a function, we will specify a rule that tells us how to associate an element of the range to an element of the domain.

EXAMPLES A.7.14.

(i) $f : \mathbb{N} \longrightarrow \mathbb{N}$, $f(n) = 2n$.
(ii) $f : \mathbb{Z} \longrightarrow \mathbb{Z}$, $f(n) = n + 6$.
(iii) $f : \mathbb{N} \longrightarrow \mathbb{Q}$, $f(n) = n$.
(iv) This is an unusual example. Let $\mathbb{Q}_+$ be the set of positive rational numbers written as fractions in base 10, and let $\mathbb{N}$ be the natural numbers but written in base 11 with the numerals being 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, d. So, for example, the integer which is written 21 in base 10 is $1d$ in base 11, and $1222_{(10)} = d11_{(11)}$. (Incidentally, we are writing the bases in base 10.) Now define a function $f : \mathbb{Q}_+ \longrightarrow \mathbb{N}$ by writing $\frac{a}{b} = \frac{a_n \ldots a_2 a_1}{b_m \ldots b_2 b_1}$, where a_i and b_i are the i-th digits of the numerator and denominator (and, of course, integers between 0 and 9). Then, set $f(\frac{a}{b}) = a_n \ldots a_2 a_1 d b_m \ldots b_2 b_1$. The fraction $\frac{a}{b}$ will always be written in lowest terms. For instance, if we take the fraction 2/3 in base 10, then $f(2/3) = 2d3$ in base 11, which is the same as the integer 355 written in base 10.
(v) Write the real numbers in terms of their decimal expansions. As usual, we do not allow a real number to end in all 9's repeating. Let $f : \mathbb{R} \longrightarrow \mathbb{N}$ be defined by: $f(x)$ equals the third digit of x after the decimal point (this is called the *Michelle* function).

EXERCISE A.7.15. Determine which of the above functions are surjective, injective, or bijective.

We mentioned above the so-called identity function, and we assign a symbol to it.

DEFINITION A.7.16. Let A be a set. We give a symbol for the identity function defined in Example A.7.1. The identity function $I_A : A \longrightarrow A$ is defined by $I_A(a) = a$ for $a \in A$.

Now suppose A and B are sets and $f : A \longrightarrow B$ is a bijection. Since each element of B comes from only one element of A under the function f, we can define a function $f^{-1} : B \longrightarrow A$ that sends every element of B back to where it came from.

DEFINITION A.7.17. Let $f : A \longrightarrow B$ be a bijection. Then the *inverse* of f is the function $f^{-1} : B \longrightarrow A$ defined as follows. If $b \in B$, then we set $f^{-1}(b) = a$ where a is the unique element of A such that $f(a) = b$.

EXERCISE A.7.18. Show that $f^{-1} \circ f = I_A$ and $f \circ f^{-1} = I_B$.

EXERCISE A.7.19. Suppose A, B, and C are sets and $f : A \longrightarrow B$ and $g : B \longrightarrow C$ are bijections. Show that $g \circ f$ is a bijection. Compute $(g \circ f)^{-1} : C \longrightarrow A$.

EXERCISE A.7.20. Given $f : A \longrightarrow B$, suppose there exist $g, h : B \longrightarrow A$ so that $f \circ g = I_B$ and $h \circ f = I_A$. Show that f is a bijection and that $g = h = f^{-1}$.

EXERCISE A.7.21. Let $\mathbb{R}_+$ be the positive real numbers and define $f : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ by $f(x) = x^2$. Show that f is a bijection from $\mathbb{R}_+$ to $\mathbb{R}_+$ and find f^{-1}. If we expand the domain and include all real numbers, what happens?

EXERCISE A.7.22. Define $f : \mathbb{N} \longrightarrow \mathbb{Z}$ by

$$f(n) = \begin{cases} \dfrac{n}{2} & \text{if } n \text{ is even,} \\ \dfrac{1-n}{2} & \text{if } n \text{ is odd.} \end{cases}$$

Show that f is a bijection.

Whether or nor a function $f : A \longrightarrow B$ is a bijection, we can take a subset $A' \subseteq A$ and consider its image $f(A') \subseteq B$. Moreover, we can take a subset B' of B and consider the *preimage* of B' in A.

DEFINITION A.7.23. Suppose A and B are sets and $f : A \longrightarrow B$ is a function. If $B' \subseteq B$, then the *preimage* of B' in A is defined as

$$f^{-1}(B') = \{a \in A \mid f(a) \in B'\}.$$

So, $f^{-1}(B')$ is everything in A that is pushed into B' by the function f. Let's make a few quick observations about the empty set.

(1) $f(\varnothing) = \varnothing$.
(2) $f^{-1}(\varnothing) = \varnothing$.
(3) More generally, if $B' \subseteq B$ and $B' \cap f(A) = \varnothing$, then $f^{-1}(B') = \varnothing$.

Take heed: given any subset $B' \subseteq B$, its preimage $f^{-1}(B')$ always exists. Despite the use of the f^{-1} notation, this has nothing to do with whether or not f has an inverse.

There are four basic results on images and preimages.

THEOREM A.7.24. *Suppose A and B are sets and $f : A \longrightarrow B$ is a function. Let $A_1, A_2 \subseteq A$ and $B_1, B_2 \subseteq B$. Then*

(i) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$;
(ii) $f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2)$;
(iii) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$;
(iv) $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.

Proof. The proof is standard. We will prove (iii). Take $x \in f^{-1}(B_1 \cup B_2)$. Then $f(x) \in B_1 \cup B_2$ so $f(x) \in B_1$ or $f(x) \in B_2$. Hence $x \in f^{-1}(B_1)$ or $x \in f^{-1}(B_2)$. That is, $x \in f^{-1}(B_1) \cup f^{-1}(B_2)$ and so $f^{-1}(B_1 \cup B_2) \subseteq f^{-1}(B_1) \cup f^{-1}(B_2)$. Actually, you can read this argument backwards to show that $f^{-1}(B_1) \cup f^{-1}(B_2) \subseteq f^{-1}(B_1 \cup B_2)$. Thus, finally, the sets are equal. This is terribly boring, but you should do (i) and (iv) to discipline yourself. On the other hand, (ii) is more interesting.

EXERCISE A.7.25. Find an example to show that equality does not necessarily hold in (ii).

EXERCISE A.7.26. Show that equality holds in (ii) of Theorem A.7.24 if f is an injection. In fact, if equality holds in (ii) for all subsets $A_1, A_2 \subseteq A$, then f is an injection.

EXERCISE A.7.27. Let A and B be sets and let $f : A \longrightarrow B$ be a function. Suppose that $\{A_i\}_{i \in I}$ is a collection of subsets of A and $\{B_j\}_{j \in J}$ is a collection of subsets of B. Show that

(i) $f(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f(A_i)$;
(ii) $f(\bigcap_{i \in I} A_i) \subseteq \bigcap_{i \in I} f(A_i)$;
(iii) $f^{-1}(\bigcup_{j \in J} B_j) = \bigcup_{j \in J} f^{-1}(B_j)$;
(iv) $f^{-1}(\bigcap_{j \in J} B_j) = \bigcap_{j \in J} f^{-1}(B_j)$.

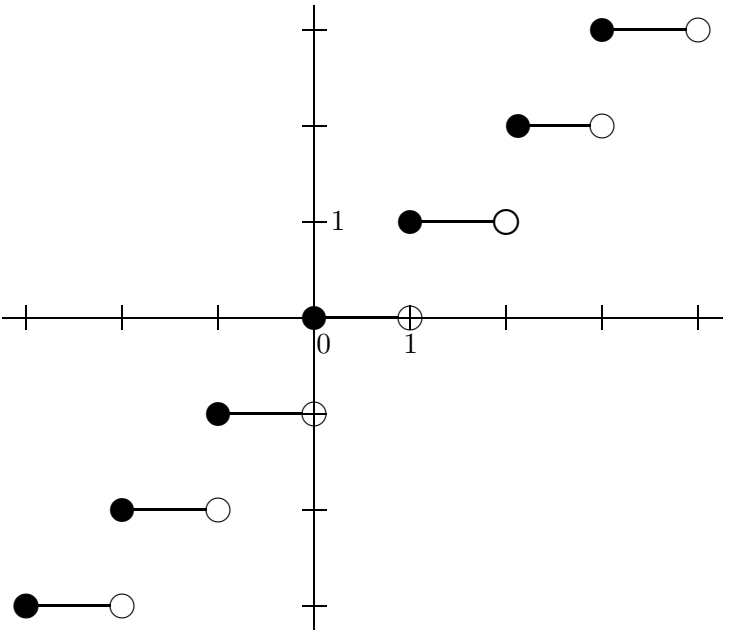
Note that, in this exercise, the number of sets in the union and intersection is not necessarily finite.

To close this section, we consider two important examples of functions.

DEFINITION A.7.28. The *greatest integer function* $[\,\cdot\,] : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by: $[x]$ equals the largest integer that is less than or equal to x.

EXAMPLE A.7.29. $[n] = n$ for $n \in \mathbb{Z}$; $[17.5] = 17$; $[\sqrt{2}] = 1$; $[\pi] = 3$; $[-e] = -3$, etc.

Here is the graph of the greatest integer function:



EXERCISE A.7.30. Express the Michelle function (Example A.7.14(v)) in terms of the greatest integer function. Graph the Michelle function.

Now we define polynomial functions. Polynomial functions are perfect examples of functions that fit into the "What is my rule?" category. Here is a polynomial function with its rule. Let $p(x)$ be the function given by $p(x) = x^2 + 2x + 1$. You can plug in numbers for x and get values for $p(x)$. For instance, $p(0) = 1$ and $p(1) = 4$, and $p(\sqrt{2}) = 3 + 2\sqrt{2}$. Let's be a little more formal here.

DEFINITION A.7.31. A *polynomial function* $p : \mathbb{R} \longrightarrow \mathbb{R}$ is a function whose rule of correspondence is given by an expression of the form

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where n is a natural number or zero and the *coefficients* $a_n, a_{n-1}, \ldots, a_1, a_0$ are in $\mathbb{R}$. Note that the function $p(x) = 0$, in which the coefficients $a_0, a_1, \ldots, a_n$ are all zero, is a polynomial function. If $p(x) \neq 0$, then there is a largest power of x (including 0) with a nonzero coefficient. This power is called the *degree* of the polynomial function. We make the convention that the degree of the polynomial function 0 is $-\infty$.

As an ordered pair, a polynomial function is written $(x, p(x))$. In the above definition, we say that p is a *polynomial function with real coefficients.* Notice that we could restrict the coefficients to be integers or rational numbers. In these cases, we could restrict the domain to be the integers or rational numbers, respectively.

Examples A.7.32.

(i) $p(x) = 0$ is a polynomial of degree $-\infty$ with coefficients in $\mathbb{Z}$.
(ii) $p(x) = 17x^2 + 2x - 7$ is a polynomial of degree 2 with coefficients in $\mathbb{Z}$.
(iii) $p(x) = (\sqrt{2} - 1)x^{83} + \pi x^{17} + \sqrt[3]{2}x^{11} + ex^7 + 6$ is a polynomial of degree 83 with coefficients in $\mathbb{R}$.
(iv) $p(x) = 17/4$ is a polynomial of degree 0 with coefficients in $\mathbb{Q}$.

You should have experience in adding and multiplying polynomial functions, so we won't go into details. Here are a couple of interesting exercises.

Exercise A.7.33. If p and q are polynomial functions with real coefficients, show that $\deg(pq) = \deg(p) + \deg(q)$. To accommodate the zero polynomial, we use the convention $-\infty + k = -\infty$ for any k.

Exercise A.7.34. If p and q are polynomial functions with real coefficients and $\deg(p) \neq \deg(q)$, show that $\deg(p+q) = \max(\deg(p), \deg(q))$. In any case, $\deg(p+q) \leq \max(\deg(p), \deg(q))$.

Exercise A.7.35. Show that the set of polynomial functions from $\mathbb{R}$ to $\mathbb{R}$ is an integral domain.

Another important type of function is that of a sequence. Sequences will occur regularly throughout the remainder of the text.

Definition A.7.36. Let X be a nonempty set. A *sequence* in X is a function $f : \mathbb{N} \longrightarrow X$.

Thus, respecting the order in $\mathbb{N}$, we write a sequence

$$f(1), f(2), \ldots, f(n), \ldots$$

or $(x_1, x_2, \ldots, x_n, \ldots)$. We will also adopt the notation $(x_n)_{n\in\mathbb{N}}$.

Remark A.7.37. It is also useful to have sequences indexed by the nonnegative integers, or even the set of all integers. So, for example, we might have $(x_0, x_1, \ldots, x_n, \ldots)$ or $(x_n)_{n\geq 0}$. If the index set is $\mathbb{Z}$, we write $(\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots)$, or $(x_n)_{n\in\mathbb{Z}}$.

Exercise A.7.38. If A_1 and A_2 are subsets of a universal set X, show that there is a bijection between the Cartesian product $A_1 \times A_2$ and the set of all functions $f : \{1, 2\} \longrightarrow X$ such that $f(1) \in A_1$ and $f(2) \in A_2$. Do the same for any finite number of subsets of X.

This exercise is the beginning of our study of the axiom of choice, which comes up later in the chapter.

A.8. Countability and Other Basic Ideas

Finally, we come to a serious discussion of infinite sets. There are great pitfalls involved in any discussion of set theory, and our basic goal is to avoid these pitfalls while still having appropriate definitions, ideas, and facts. In

analysis, most of the sets we deal with are infinite. In fact, most of them contain the integers in one way or another. Moreover, any discussion of continuity and change involves infinite sets. So, of course, in our usual perverse manner, we define finite sets first.

DEFINITION A.8.1. A set A is *finite* if A is empty or there exists $n \in \mathbb{N}$ such that there is a bijection $f : A \to \{1, 2, \ldots, n\}$, where $\{1, 2, \ldots, n\}$ is the set of all natural numbers less than or equal to n. In this case, we say A has n elements.

EXERCISE A.8.2. If A is a finite set and B is a subset of A, show that B is a finite set. In addition, show that if B is a proper subset, then the number of elements in B is less than the number of elements in A.

There is a natural and useful characteristic of finite sets:

THEOREM A.8.3. *If A is a finite set and B is a proper subset of A, then there is no bijection between B and A.*

Proof. Suppose A has n elements and B has m elements with $m < n$. Then the pigeonhole principle tells us that, for any function from A to B, there is an element of B that is the image of at least two different elements of A.

EXERCISE A.8.4. Show that the following are finite sets:

(i) the English alphabet,
(ii) the set of all possible twelve-letter words made up of letters from the English alphabet,
(iii) the set of all subsets of a finite set.

This approach to things makes the definition of infinite sets quite simple:

DEFINITION A.8.5. An *infinite* set is a set that is not finite.

The notion of cardinality of a set is very important. Of course, most authors don't define cardinality. Instead, they say what it means for two sets to have the same cardinal number. This will do for our purposes.

DEFINITION A.8.6. The *cardinal number* of a nonempty finite set A is the number of elements in A; that is, the cardinal number of A is the natural number n if there is a bijection between A and $\{k \in \mathbb{N} \mid 1 \leq k \leq n\}$. The cardinal number of the empty set is 0.

DEFINITION A.8.7. A set A has *cardinality* $\aleph_0$ (pronounced "aleph null" or "aleph naught") if it can be put in one-to-one correspondence with $\mathbb{N}$, that is, there is a bijection between the set and $\mathbb{N}$. In general, two sets have the same cardinality if they can be put in one-to-one correspondence with each other.

EXAMPLE A.8.8. The set $\mathbb{N}$ has cardinality $\aleph_0$ (this should not come as a surprise).

Although we will not see one for a while, be assured that there are infinite sets with cardinality other than $\aleph_0$.

EXAMPLE A.8.9. The set $\mathbb{N} \cup \{0\}$ has cardinality $\aleph_0$ because the function $f : \mathbb{N} \cup \{0\} \to \mathbb{N}$ given by $f(n) = n + 1$ is a bijection.

EXAMPLE A.8.10. The set $\mathbb{Z}$ has cardinality $\aleph_0$ because the function $f : \mathbb{Z} \to \mathbb{N}$ given by

$$f(z) = \begin{cases} 2z + 2 & \text{if } z \geq 0, \\ -2z - 1 & \text{if } z < 0 \end{cases}$$

is a bijection.

There is a very useful theorem that asserts the existence of a one-to-one correspondence between two sets. This relieves us of the burden of constructing a bijection between two sets to show that they have the same cardinality.

THEOREM A.8.11 **(Schröder-Bernstein)**. *If A and B are sets and if there exist injections $f : A \to B$ and $g : B \to A$, then there exists a bijection between A and B.*

Proof. First, we divide A into three disjoint subsets. For each $x \in A$, consider the list of elements

$$S_x = \{x, g^{-1}(x), f^{-1} \circ g^{-1}(x), g^{-1} \circ f^{-1} \circ g^{-1}(x), \dots\}.$$

The elements of this sequence are called *predecessors* of x. Notice that in S_x, we start with $x \in A$. Then $g^{-1}(x) \in B$ if $g^{-1}(x)$ exists (x may not be in the image of g). For each $x \in A$, exactly one of the three following possibilities occurs.

(1) The list S_x is infinite.
(2) The last term in the list is an element of A. That is, the last term is $y = x$ or is of the form $y = f^{-1} \circ g^{-1} \circ \cdots \circ g^{-1}(x)$ and $g^{-1}(y)$ does not exist (i.e., y is not in the image of g). In this case, we say that S_x *stops in* A.
(3) The last term in the list is an element of B. That is, the last term is $z = g^{-1}(x)$ or is of the form $z = g^{-1} \circ f^{-1} \circ \cdots \circ g^{-1}(x)$ and $f^{-1}(z)$ does not exist (i.e., z is not in the image of f). In this case, we say that S_x *stops in* B.

Let the corresponding subsets of A be denoted by A_1, A_2, A_3. Similarly, define the corresponding subsets of B. That is,

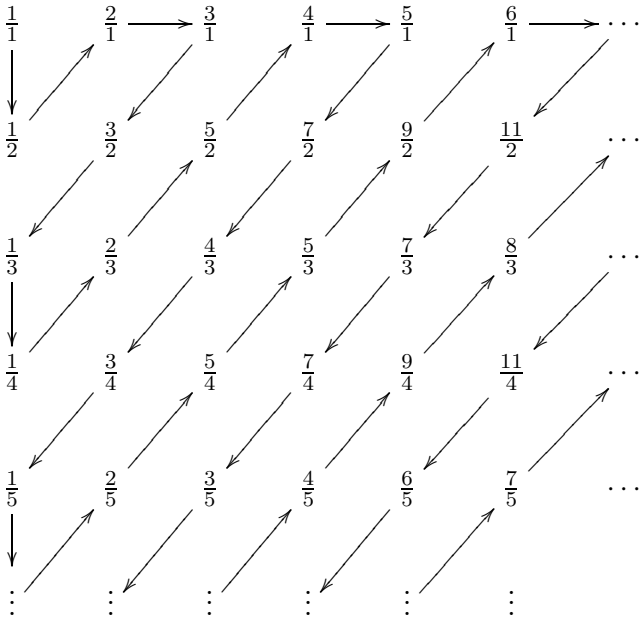
$$\begin{aligned} B_1 &= \{y \in B \mid y \text{ has infinitely many predecessors}\}, \\ B_2 &= \{y \in B \mid \text{the predecessors of } y \text{ stop in } A\}, \text{ and} \\ B_3 &= \{y \in B \mid \text{the predecessors of } y \text{ stop in } B\}. \end{aligned}$$

Now observe that $f : A_1 \longrightarrow B_1$, $g : B_1 \longrightarrow A_1$ are both bijections. Also, $f : A_2 \longrightarrow B_2$ and $g : B_3 \longrightarrow A_3$ are bijections.

EXERCISE A.8.12. Suppose A, B, and C are subsets of a set X such that $A \subseteq B \subseteq C$. Show that if A and C have the same cardinality, then A and B have the same cardinality.

EXAMPLE A.8.13. The set $\mathbb{Q}_+$ has cardinality $\aleph_0$ (recall that $\mathbb{Q}_+$ denotes the positive rational numbers). Here are three proofs:

(1) This is a very common and very sloppy proof. However, the underlying idea will stand us in good stead.



To find a bijection between $\mathbb{N}$ and $\mathbb{Q}_+$, we write all the positive fractions in a grid, with all fractions with denominator 1 in the first row, all fractions with denominator 2 in the second row, all fractions with denominator 3 in the third row, etc. Now go through row-by-row and throw out all the fractions that aren't written in lowest terms. Then, starting at the upper left-hand corner, trace a path through all the remaining numbers as above.

We can count along the path we drew, assigning a natural number to each fraction. So $\frac{1}{1} \to 1$, $\frac{1}{2} \to 2$, $\frac{2}{1} \to 3$, $\frac{3}{1} \to 4$, $\frac{3}{2} \to 5$, etc. This is a bijection. Therefore, $\mathbb{Q}_+$ has cardinality $\aleph_0$. Although this is a very common proof, the formula for the bijection is not at all obvious. It is very difficult to see, for example, which rational number corresponds to 1,000,000.

(2) In this proof, we'll make use of the Schröder-Bernstein theorem. It is easy to inject $\mathbb{N}$ into $\mathbb{Q}_+$: simply send n to n. The injection from $\mathbb{Q}_+$ to $\mathbb{N}$ will be the one we used in Example A.7.14(iv) (where $\frac{a}{b}$ is

sent to $adb_{(11)}$). Each number which is the image of a fraction has one and only one d in it, so it is easy to see which fraction is represented by a given integer. According to the Schröder-Bernstein theorem, two injections make a bijection, so $\mathbb{Q}_+$ has cardinality $\aleph_0$.

(3) Write each positive fraction in lowest terms and factor the numerator and denominator into primes, so that $\frac{p}{q} = \frac{p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}}{q_1^{\beta_1} q_2^{\beta_2} \cdots q_m^{\beta_m}}$, with $p_i \neq q_j$. If by chance p or q is 1 and can't be factored, write it as 1^1. Then let $f : \mathbb{Q}_+ \to \mathbb{N}$ be defined by

$$f\left(\frac{p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}}{q_1^{\beta_1} q_2^{\beta_2} \cdots q_m^{\beta_m}}\right) = p_1^{2\alpha_1} p_2^{2\alpha_2} \cdots p_n^{2\alpha_n} q_1^{2\beta_1 - 1} q_2^{2\beta_2 - 1} \cdots q_m^{2\beta_m - 1}.$$

In particular, note that if $a \in \mathbb{Q}_+$ is an integer, then $f(a) = a^2$.

EXERCISE A.8.14. Verify that f is a bijection.

EXERCISE A.8.15. Suppose that $N = 10^k$ for some integer k. Find $p/q \in \mathbb{Q}_+$ such that $f(p/q) = N$.

EXERCISE A.8.16. Use any one of the above three proofs to show that $\mathbb{Q}$ has cardinality $\aleph_0$.

EXERCISE A.8.17. Show that the natural numbers form an infinite set.

EXERCISE A.8.18. Show that any set that has the same cardinal number as $\mathbb{N}$ is an infinite set.

Note: A set is called *countably infinite* if it has cardinality $\aleph_0$ (that is, if it is in one-to-one correspondence with the natural numbers). The term "countable" (or sometimes called "*denumerable*") is used in several ways. Many people use it to refer to infinite sets that are in one-to-one correspondence with $\mathbb{N}$, while others include finite sets when they say countable. This is not something to get disturbed about. We will say that a set is *countable* if it is finite, or countably infinite.

EXERCISE A.8.19. Show that a subset of a countable set is countable.

EXERCISE A.8.20. Show that the set of all polynomial functions with integer coefficients is a countable set.

THEOREM A.8.21. *If A is an infinite set, then A has a countably infinite subset.*

Proof. Take any infinite set A and choose an element a_1 in A. Let $A_1 = A \setminus \{a_1\}$. By the definition of infinite set, A_1 is infinite. So we choose a_2 in A_1 and define $A_2 = A \setminus \{a_1, a_2\}$. Since A is not finite, we can continue

to choose elements. Thus, if we have chosen $a_1, \ldots, a_n$, we consider $A_n = A \setminus \{a_1, \ldots, a_n\}$. Since A is infinite, we can choose an element a_{n+1} in A_n. Continuing inductively, we obtain our desired countable subset. Note that this countable set may be all of A.

REMARK A.8.22. There is some discussion among mathematicians as to whether the preceding proof involves the axiom of choice. The axiom of choice in its fullest form will be discussed below. However, one can make the argument that it requires some sort of choice mechanism to pick an element from a nonempty set. The technique that we use in the proof of Theorem A.8.21 is sometimes referred to as "the countable axiom of choice".

We could pursue an alternate definition of an infinite set. In fact, we could define infinite sets first and then say that a finite set is a set that is not infinite.

REDEFINITION A.8.23. A set is *infinite* if there is a bijection between the set and one of its proper subsets.

REDEFINITION A.8.24. A *finite* set is a set that is not infinite.

To show the equivalence of the two definitions, recall that in Theorem A.8.3 we showed that there is no bijection between a finite set and any of its proper subsets. This means that if a set is infinite by our new definition, it is not finite (hence, infinite) by the old definition too. Next, let's show that any set that is infinite by the old definition is bijective with one of its proper subsets.

Proof. Say A is an infinite set and $B \subseteq A$ is countably infinite. Then we can write $B = \{b_1, b_2, \ldots, b_n, \ldots\}$. Now define $f : A \to A \setminus \{b_1\}$ as follows: for $a \in A \setminus B$, $f(a) = a$, and for $b_i \in B$, $f(b_i) = b_{i+1}$. Thus f is a bijection between A and $A \setminus \{b_1\}$. Therefore, our definitions are equivalent.

We now turn to operations involving infinite sets.

FACTS A.8.25.

(1) If A_1 and A_2 are countable sets, then $A_1 \cup A_2$ is a countable set.
(2) If A_1, A_2, ..., A_n are countable sets, then $\bigcup_{j=1}^{n} A_j$ is a countable set.
(3) Let $\{A_j\}_{j \in \mathbb{N}}$ be a countable collection of countable sets. Then $\bigcup_{j \in \mathbb{N}} A_j$ is a countable set.

Proof. We prove (3) only. You can prove the other two (or deduce them from (3)).

Write $A_j = \{a_{j,1}, a_{j,2}, \ldots, a_{j,n}, \ldots\}$. We use the diagonal process as in Example A.8.13. Simply write

$$\begin{aligned} A_1 &: a_{1,1}, a_{1,2}, \ldots, a_{1,n}, \ldots, \\ A_2 &: a_{2,1}, a_{2,2}, \ldots, a_{2,n}, \ldots, \\ &\vdots \\ A_m &: a_{m,1}, a_{m,2}, \ldots, a_{m,n}, \ldots, \\ &\vdots \end{aligned}$$

Now count diagonally, ignoring repetitions.

Let's take a look now at Cartesian products. It is clear from the ideas presented above that if A_1 and A_2 are countable, then $A_1 \times A_2$ is countable.

EXERCISE A.8.26.

(i) Show that if A_1, A_2, ..., A_n are countable, then $A_1 \times A_2 \times \cdots \times A_n$ is countable.

(ii) What can you say about the countable Cartesian product of countable sets?

Next we look at the power set $\mathscr{P}(A)$ for any set A.

THEOREM A.8.27. *If A is any set (including the empty set), there is no bijection between A and $\mathscr{P}(A)$.*

Proof. This is clear if A is the empty set. Suppose that there is a bijection between A and $\mathscr{P}(A)$. If $a \in A$, let P_a be the subset of A associated with it. Now consider the set $B = \{a \mid a \notin P_a\}$. The set B must be associated to some element of A, which we creatively call b, so that $B = P_b$. Is b in B? For b to be in B, we must have that $b \notin P_b$. But $B = P_b$, so therefore b is not in B. But then $b \in P_b$, which means that b is in B. This is a contradiction. Therefore, there is no bijection between A and $\mathscr{P}(A)$.

DEFINITION A.8.28. If A is a countably infinite set, then the cardinality of $\mathscr{P}(A)$ is denoted by $\mathbf{c}$.

EXERCISE A.8.29. Show that the definition of the cardinal number $\mathbf{c}$ does not depend on the choice of the countably infinite set A. That is, if A and B are countably infinite sets, then there is a bijection between $\mathscr{P}(A)$ and $\mathscr{P}(B)$.

REMARK A.8.30. At this point, we observe that if A is a countably infinite set, $A = \{a_1, a_2, \ldots, a_n, \ldots\}$, then $\mathscr{P}(A)$ is in one-to-one correspondence with the set of all functions from A to the set $\{0, 1\}$. This correspondence is defined as follows. If B is a subset of A, then we define the map $f_B : A \to \{0, 1\}$ by $f_B(a_j) = 1$ if a_j is in B, 0 if a_j is not in B. In accordance with the notation of Theorem A.4.13, we will write $\mathbf{c} = 2^{\aleph_0}$. Observe that $f_B(a)$ can be viewed as a binary expansion of a real number between 0 and 1.

EXERCISE A.8.31. Suppose that A is a nonempty set. Show that $\mathscr{P}(A)$ is in one-to-one correspondence with the set of all functions from A to $\{0,1\}$.

One of the most important sets of numbers that we deal with in this book is the collection of real numbers $\mathbb{R}$. In Chapter 3, we will go through the formal construction of the real numbers from the rational numbers. For the present discussion, we can just consider the set of real numbers to be the set of all terminating or infinite decimals with the convention that no decimal expansion can terminate in all 9's. There are two things to show about the reals. The first is the proof due to Cantor that the reals are uncountable, and the second is that the cardinality of the real numbers is in fact $\mathbf{c}$.

THEOREM A.8.32. *The set of all real numbers between* 0 *and* 1 *is not countable.*

Proof. We first note that the decimal expansion of a real number is unique, since it does not end in all 9's. In this case, we always round up the digit that occurs before the sequence of 9's. To prove that this set is not countable, we assume that it is and list the real numbers between 0 and 1 vertically:

$$\begin{aligned} a_1 &= 0.a_{1,1}a_{1,2}\dots a_{1,n}\dots, \\ a_2 &= 0.a_{2,1}a_{2,2}\dots a_{2,n}\dots, \\ &\vdots \\ a_m &= 0.a_{m,1}a_{m,2}\dots a_{m,n}\dots, \\ &\vdots \end{aligned}$$

We now proceed using a process similar to the one used in the proof of Theorem A.8.27 to produce a real number between 0 and 1 that is not on our list. We construct a number $b = 0.b_1b_2\dots b_n\dots$ by proceeding diagonally down the list as follows: if $a_{1,1} = 1$, take $b_1 = 2$, and if $a_{1,1} \neq 1$, take $b_1 = 1$. Next, if $a_{2,2} = 1$, take $b_2 = 2$, and if $a_{2,2} \neq 1$, take $b_2 = 1$. Continuing this process, we see that the decimal $b = 0.b_1b_2\dots b_n\dots$ is a real number, since it does not end in all 9's, but it cannot be on our list, since it differs from each number we list in at least one digit. Consequently, the real numbers between 0 and 1 are not countable.

THEOREM A.8.33. *The cardinality of the real numbers between* 0 *and* 1 *is* $\mathbf{c} = 2^{\aleph_0}$.

Proof. To write down an exact bijection between $\mathscr{P}(\mathbb{N})$ and the real numbers between 0 and 1 requires some care. The standard way to do this is to write the binary expansions of all real numbers between 0 and 1 in such a way that no expansion terminates in all 1's. In considering the corresponding subsets of $\mathbb{N}$, we first remove two specific subsets of $\mathscr{P}(\mathbb{N})$. We remove the two collections $A_f = \{C \in \mathscr{P}(\mathbb{N}) \mid C \text{ is finite}\}$ and $A_{cf} = \{D \in \mathscr{P}(\mathbb{N}) \mid {}^cD \text{ is finite}\}$. The collection $\mathscr{P}(\mathbb{N}) \setminus (A_f \cup A_{cf})$

is in one-to-one correspondence with all binary expansions which have an infinite number of 1's but do not terminate in all 1's. We get the required bijection by Remark A.8.30.

We can place A_f into one-to-one correspondence with the set of all finite binary expansions with 0 in the first place, and A_{cf} into one-to-one correspondence with the set of all finite binary expansions with 1 in the first place.

EXERCISE A.8.34. Write down these last two bijections explicitly.

EXERCISE A.8.35.

(i) Prove that the countable union of sets of cardinality $\mathbf{c}$ again has cardinality $\mathbf{c}$.
(ii) Prove that the set of all real numbers has cardinality $\mathbf{c}$.
(iii) Prove that the set of irrational numbers in $\mathbb{R}$ has cardinality $\mathbf{c}$.

How big do cardinal numbers get? For instance, the power set of $\mathbb{R}$ is "bigger than" $\mathbb{R}$. In fact, the power set of $\mathbb{R}$ can be identified with the set of all maps from $\mathbb{R}$ into $\{0,1\}$ just as we did above for the power set of $\mathbb{N}$. Thus, we have $\#(\mathscr{P}(\mathbb{R})) = 2^{\mathbf{c}}$. We sometimes denote $2^{\mathbf{c}}$ by $\mathbf{f}$.

The following theorem is interesting and useful.

THEOREM A.8.36. *There is a bijection between the unit interval and the unit square.*

Proof. Let $I = [0,1] = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$ and $I^2 = [0,1] \times [0,1]$. This seems like a great time to use the Schröder-Bernstein theorem. The function $f : I \to I^2$ defined by $f(x) = (x, 0)$ is an injection. Define the function $g : I^2 \longrightarrow I$ by the rule $g((a_0.a_1a_2 \ldots a_n \ldots, b_0.b_1b_2 \ldots b_n \ldots)) = (0.a_0b_0a_1b_1a_2b_2 \ldots a_nb_n \ldots)$, where $a_0.a_1a_2 \ldots a_n \ldots$ and $b_0.b_1b_2 \ldots b_n \ldots$ are decimal expansions of the coordinates of any point in I^2 (of course, the decimal expansion is prohibited from ending in all 9's). The function $g : I^2 \to I$ is an injection. Therefore, there is a bijection between I and I^2.

A.9. The Axiom of Choice

DEFINITION A.9.1. A *partially ordered set* is a set X with a relation $\leq$ that is reflexive, transitive, and antisymmetric (that means that if $a \leq b$ and $b \leq a$, then $a = b$). A *totally ordered set* is a partially ordered set with the additional property that, for any two elements $a, b \in X$, either $a \leq b$ or $b \leq a$. A *well-ordered set* is a totally ordered set in which any nonempty subset has a least element.

EXAMPLE A.9.2.

(1) $(\mathbb{N}, \leq)$ is a totally ordered set, as are $(\mathbb{Z}, \leq)$, $(\mathbb{Q}, \leq)$, and $(\mathbb{R}, \leq)$.
(2) Let X be a set, and let $\mathscr{P}(X)$ be the collection of all subsets of X. Then $(\mathscr{P}(X), \subseteq)$ is a partially ordered set.

DEFINITION A.9.3. Let Y be a subset of a partially ordered set X. An *upper bound* for Y is an element $a \in X$ such that $y \leq a$ for all $y \in Y$. A *least upper bound* for Y is an element $b \in X$ such that b is an upper bound for Y, and if a is an upper bound for Y, then $b \leq a$. The least upper bound is sometimes abbreviated lub and is also denoted as sup (supremum). You can figure out what a *lower bound* and *greatest lower bound* (glb) are. The greatest lower bound is also denoted by inf (infimum).

Observe that a subset of a partially ordered set may fail to have an upper bound or a lower bound.

EXERCISE A.9.4. If a subset Y of a partially ordered set X has an upper bound, determine whether or not Y must have a least upper bound. If Y has a least upper bound, determine whether or not this least upper bound is unique.

DEFINITION A.9.5. In a partially ordered set, an element b is *maximal* if $a \geq b$ implies $a = b$.

We turn now to one of the major topics of this chapter, the axiom of choice, and various logically equivalent statements. For many years, there has been considerable discussion among mathematicians about the use of the axiom of choice and the seemingly contradictory results that come along with it. We find it indispensable in obtaining a number of results in mathematics.

The Axiom of Choice A.9.6. Given a collection $\mathcal{C}$ of sets that does not include the empty set, there exists a function $\phi : \mathcal{C} \to \bigcup_{C \in \mathcal{C}} C$ with the property that $\forall A \in \mathcal{C}$, $\phi(A) \in A$.

Another way of looking at this is as follows. Suppose $\{A_i\}_{i \in I}$ is a collection of nonempty sets indexed by an index set I. A *choice function* is then defined as a map $\phi : I \to \bigcup_{i \in I} A_i$ such that $\phi(i) \in A_i$. The axiom of choice can then be rephrased.

The Axiom of Choice A.9.7. For every collection of nonempty sets there exists a choice function.

The axiom of choice is equivalent to a number of other very useful statements which are not at all obvious. Here they are, in no particular order.

Let X be a partially ordered set. The collection $\mathscr{P}(X)$ can be partially ordered by inclusion; see Example A.9.2. This partial ordering on $\mathscr{P}(X)$ is used in some of the statements below.

Hausdorff Maximality Principle A.9.8. Every partially ordered set X contains a totally ordered subset that is maximal with respect to the ordering on $\mathscr{P}(X)$.

Zorn's Lemma A.9.9. If a nonempty partially ordered set has the property that every nonempty totally ordered subset has an upper bound, then the partially ordered set has a maximal element.

Well-Ordering Principle A.9.10. Every set can be well-ordered.

This is quite a different statement from the well-ordering principle for the integers.

The following lemma is slightly complicated, but it will allow us to prove the equivalence of the above statements with little trouble.

LEMMA A.9.11. *Suppose that $(X, \leq)$ is a nonempty partially ordered set such that every nonempty totally ordered subset has a least upper bound. If $f : X \to X$ is such that $f(x) \geq x$ for all $x \in X$, then there is some $w \in X$ such that $f(w) = w$.*

Proof. First we reduce to the case when X contains a least element; call it b. In fact, if X is nonempty, choose any $b \in X$ and replace X by $X' = \{x \in X \mid x \geq b\}$. It is clear that X' is stable under f (that is, $f(X') \subseteq X'$) and that it has the same properties as X. We call a subset Y of X "admissible" if

(1) $b \in Y$,
(2) $f(Y) \subseteq Y$,
(3) every lub of a totally ordered subset of Y belongs to Y.

X is certainly admissible, and the intersection of any family of admissible sets is admissible. Let W be the intersection of all admissible sets. The set $\{x \in X \mid b \leq x\}$ is admissible, so if $y \in W$, then $b \leq y$.

We will now construct a totally ordered subset of W with the property that its least upper bound is a fixed point of f. Consider the set $P = \{x \in W \mid \text{ if } y \in W \text{ and } y < x, \text{ then } f(y) \leq x\}$. Note that P is nonempty since $b \in P$. First we show that any element of P can be compared to any element of W and hence P is totally ordered.

Now fix an $x \in P$ and define $A_x = \{z \in W \mid z \leq x \text{ or } z \geq f(x)\}$. We would like to show that A_x is admissible.

(1) Obviously, $b \in A_x$ since $b \leq x$.
(2) Suppose $z \in A_x$. There are three possibilities. If $z < x$, $f(z) \leq x$ by the conditions of P, so $f(z) \in A_x$. If $z = x$, $f(z) = f(x) \geq f(x)$ so $f(z) \in A_x$. If $z \geq f(x)$, then $f(z) \geq z \geq f(x)$ so $f(z) \in A_x$.
(3) Finally, let Y be a totally ordered nonempty subset of A_x, and let y_0 be the lub of Y in X. Then $y_0 \in W$, since W is admissible. If $z \leq x$ for all $z \in Y$, then $y_0 \leq x$ and hence $y_0 \in A_x$. Otherwise, $z \geq f(x)$ for some $z \in Y$, which implies $y_0 \geq f(x)$, so $y_0 \in A_x$.

Thus, A_x is admissible.

Since A_x is an admissible subset of W, $A_x = W$. Put another way, if $x \in P$ and $z \in W$, then either $z \leq x$ or $z \geq f(x) \geq x$, and thus P is totally ordered. Therefore P has a least upper bound; call it x_0. Again $x_0 \in W$ and $f(x_0) \in W$ because W is admissible. We will now show $f(x_0) = x_0$. First we claim $x_0 \in P$. Indeed, if $y \in W$ and $y < x_0$, then there exists $x \in P$ with $y < x \leq x_0$, whence $f(y) \leq x \leq x_0$. Let $y \in W$ and suppose

$y < f(x_0)$. As we saw above, $A_{x_0} = W$, so we have $y \leq x_0$. If $y = x_0$, then $f(y) = f(x_0) \leq f(x_0)$. If $y < x_0$, then $f(y) \leq x_0 \leq f(x_0)$. In either case, we find $f(x_0) \in P$. Hence $f(x_0) \leq x_0 \leq f(x_0)$.

Whew!

THEOREM A.9.12. (1) *The axiom of choice,* (2) *the Hausdorff maximality principle,* (3) *Zorn's lemma, and* (4) *the well-ordering principle are all equivalent.*

Proof. We will show that (1) implies (2), which implies (3), which implies (4), which implies (1), and then we will be done.

(1) $\Rightarrow$ (2) Take a nonempty partially ordered set $(E, \leq)$. Make $\mathcal{E}$, the family of totally ordered subsets of E, into a partially ordered set under inclusion. We wish to show that $\mathcal{E}$ has a maximal element (i.e., an element that is not smaller than any other element). So we will assume the opposite and reach a contradiction by applying Lemma A.9.11. We must first check to see if the lemma is applicable: suppose $\mathcal{F}$ is a totally ordered subset of $\mathcal{E}$. Then it has a least upper bound, namely $\bigcup_{F \in \mathcal{F}} F$. Now, for a given $e \in \mathcal{E}$, let $S_e = \{x \in \mathcal{E} \mid e \subseteq x, e \neq x\}$. Then S_e can never be the empty set because that would mean that e is maximal. So we apply the axiom of choice by defining a function $f : \{S_e \mid e \in \mathcal{E}\} \to \mathcal{E}$ with the property that $f(S_e) \in S_e$. Now define $g : \mathcal{E} \to \mathcal{E}$ by $g(e) = f(S_e)$. This gives us that $e \subsetneq g(e)$ for all $e \in \mathcal{E}$, contradicting the lemma.

(2) $\Rightarrow$ (3) Again, consider a partially ordered set $(E, \leq)$. Now let x be an upper bound for E_0, a maximal totally ordered subset of E. Suppose that there is some $y \in E$ such that $y > x$. Then $E_0 \cup \{y\}$ is a totally ordered set containing E_0, contradicting our assumption of maximality.

EXERCISE A.9.13. Now you finish the proof. Show that Zorn's lemma implies the well-ordering principle and that the well-ordering principle implies the axiom of choice.

A.10. Independent Projects

A.10.1. Basic Number Theory.

The following statements present a number of facts about elementary number theory. Prove all of these. If you don't understand some of the words, find a number theory book and look them up. Most of these facts will be used in Chapter 3 of this book when we discuss p-adic numbers. The notation $a \equiv b \pmod{c}$ (pronounced "a is *congruent* to b *modulo* c") means that $c|(a - b)$.

(1) The division algorithm: if $a, b \in \mathbb{Z}$ and $b \neq 0$, then there is a unique pair $q, r \in \mathbb{Z}$ with $a = qb + r$ and $0 \leq r < |b|$.

(2) If M is a subset of $\mathbb{Z}$ which is closed under subtraction and contains a nonzero element, then $M = \{np \mid n \in \mathbb{Z}\}$, where p is the least positive element of M.
(3) If the greatest common divisor of a and b is d, which is denoted by $d = (a, b)$, then there exist $s, t \in \mathbb{Z}$ such that $d = sa + tb$.
(4) Euclid's lemma: if p is prime and $p|ab$, then $p|a$ or $p|b$.
(5) If $(a, c) = 1$ and $c|ab$, then $c|b$.
(6) If $(a, c) = 1$, $a|m$, and $c|m$, then $ac|m$.
(7) If $a > 0$, then $(ab, ac) = a(b, c)$.
(8) The integers $\mathbb{Z}$ have unique factorization, that is, if n is an integer greater than or equal to 2, then there exist unique distinct primes $p_1, p_2, \ldots, p_k$, with $p_1 < p_2 < \cdots < p_k$, and exponents $\alpha_1, \alpha_2, \ldots, \alpha_k$ greater than or equal to 1 such that $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$.
(9) If n is a positive integer greater than or equal to 2 with unique factorization $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$, then the number of positive divisors of n is

$$(\alpha_1 + 1)(\alpha_2 + 1) \cdots (\alpha_k + 1).$$

(10) If $a \equiv b \pmod{m}$, then $-a \equiv -b \pmod{m}$, $a + x \equiv b + x \pmod{m}$, and $ax \equiv bx \pmod{m}$ for every $x \in \mathbb{Z}$.
(11) If $(c, m) = 1$ and $ca \equiv cb \pmod{m}$, then $a \equiv b \pmod{m}$.
(12) If $(c, m) = 1$, then $cx \equiv b \pmod{m}$ has a unique solution x modulo m.
(13) If p is prime and $c \not\equiv 0 \pmod{p}$, then $cx \equiv b \pmod{p}$ has a unique solution x modulo p.
(14) If $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $a + c \equiv b + d \pmod{m}$ and $ac \equiv bd \pmod{m}$.
(15) If $a, b, c \in \mathbb{Z}$ and $d = (a, b)$, then $ax + by = c$ has a solution in integers x, y if and only if $d|c$.
(16) If $[a, b]$ is the least common multiple of a and b, then $m[a, b] = [ma, mb]$ when $m > 0$.
(17) If $ca \equiv cb \pmod{m}$ and $d = (c, m)$, then $a \equiv b \pmod{\frac{m}{d}}$.
(18) If $m, a, b \in \mathbb{Z}$, the congruence $ax \equiv b \pmod{m}$ is solvable if and only if $(a, m)|b$. There are exactly (a, m) solutions that are distinct modulo m.
(19) If $a, b, s, t \in \mathbb{Z}$ are such that $sa + tb = 1$, then $(a, b) = 1$.

Now suppose that P is the set of integers between 1 and $m - 1$, inclusive, which are relatively prime to m. A *reduced residue system* modulo m is a set of integers such that each of the integers in P is congruent modulo m to exactly one of the elements in this set.

(20) The number of elements in a reduced residue system modulo m is independent of the representatives chosen.
(21) If p is a prime and ϕ denotes Euler's ϕ function (where $\phi(a)$ is the number of integers between 0 and a that are relatively prime to a), then $\phi(p^n) = p^n - p^{n-1} = p^n(1 - \frac{1}{p})$.
(22) The number of elements in a reduced residue system modulo m is $\phi(m)$.

(23) If $a_1, \ldots, a_{\phi(m)}$ is a reduced residue system modulo m and $(\kappa, m) = 1$, then $\kappa a_1, \ldots, \kappa a_{\phi(m)}$ is a reduced residue system modulo m.
(24) If m is a positive integer and $(\kappa, m) = 1$, then $\kappa^{\phi(m)} \equiv 1 \pmod{m}$.
(25) If $d_1, \ldots, d_k$ are the positive divisors of n, then $\sum_{i=1}^{k} \phi(d_i) = n$.

A.10.2. The Complete Independence of Axiom Systems.

The rules of arithmetic and order which characterize the integers are also known as *axioms*. In general, an axiom is an assumption or rule that we accept without proof. In fact, we made cancellation for multiplication an axiom for an integral domain precisely because we could not prove it from the other axioms of addition and multiplication. Note, however, that in the integers we could prove multiplicative cancellation once we stated the axioms for order. A group of axioms is called an *axiom system*.

There are a number of questions we can ask about a given axiom system S. First, is it *consistent*? That is, is there a *model* for S? For instance, the axioms for a field have a model, namely the integers modulo 2.

The next question we could ask is whether any axiom A in S is *independent*. What we mean is, "Could we replace A with its negation, $\overline{A}$, and still have a consistent system?" (Symbolically, we would represent our new axiom system as $(S - A) + \overline{A}$.)

EXERCISE A.10.1. What is the negation of "P(b), for all $b \in B$"? What about the negation of "P(b), for some $b \in B$"?

Consider the axioms for an equivalence relation. Clearly, they are consistent, because the relation of equality on any set is a model. Suppose we remove axiom (1) (reflexivity) and replace it by its negation, ($\bar{1}$), by which we mean that $(a, a) \notin R$ for some a in A. Can we come up with a model for ($\bar{1}$), (2), and (3)? To do this, pick $a \in A$ and do not include $(a, a) \in \mathcal{R}$. For the remaining elements you can fix it up so that (2) and (3) work.

EXERCISE A.10.2. State ($\bar{2}$) and ($\bar{3}$) for the equivalence relation axioms (nonsymmetry and nontransitivity). How is nonsymmetry different from antisymmetry?

The axiom system S is called *independent* if each of its axioms is independent. It is called *completely independent* if, for any subset S_1 of S, the system $(S - S_1) + \overline{S_1}$ is consistent.

EXERCISE A.10.3. Show that the axioms for an equivalence relation are completely independent. You can do this by providing models for {(1), (2), (3)}; {($\bar{1}$), (2), (3)}; {(1), ($\bar{2}$), (3)}; {(1), (2), ($\bar{3}$)}; {($\bar{1}$), ($\bar{2}$), (3)}; {($\bar{1}$), (2), ($\bar{3}$)}; {(1), ($\bar{2}$), ($\bar{3}$)}; and {($\bar{1}$), ($\bar{2}$), ($\bar{3}$)}. Your models can, but need not, be based on relations you've seen before. Or, you could invent a relation on a set which satisfies the necessary axioms. For example, a model for {(1), ($\bar{2}$), ($\bar{3}$)} could be the relation on $\{a, b, c\}$ defined by $R = \{(a, a), (b, b), (c, c), (a, b), (b, c)\}$.

The notion of complete independence will arise again when we discuss groups in the project in Section B.6.1.

A.10.3. Ordered Integral Domains.

This project is designed to show that any ordered integral domain contains a copy of the integers. Thus, in particular, any ordered field such as the rationals or real numbers contains a copy of the integers. Let $(R, +, \cdot, <)$ be an ordered integral domain.

DEFINITION A.10.4. An inductive set in R is a subset S of R such that:

(a) $1 \in S$;
(b) if $x \in S$, then $x + 1 \in S$.

EXAMPLE A.10.5.

(i) R is an inductive subset of R.
(ii) $S = \{x \in R \mid x \geq 1\}$ is an inductive subset of R.

Now define N to be the intersection of all the inductive subsets of R. It is clear that N is an inductive subset of R. Of course, N is supposed to be the natural numbers. Since all of the axioms for a commutative ring with 1, as well as the order axioms, hold in R, we can use them freely in N. The following facts are easy to prove, so prove them.

EXERCISE A.10.6.

(i) Suppose that S is a nonempty subset of N such that $1 \in S$ and if $x \in S$, then $x + 1 \in S$. Show that $S = N$.
(ii) Show that N is closed under addition.
(iii) Show that N is closed under multiplication. Hint: Fix $x \in N$ and look at the set $M_x = \{y \in N \mid xy \in N\}$. Then $1 \in M_x$. If $y \in M_x$, then $x(y+1) = xy+x$, and xy is in N by the induction hypothesis. Since N is closed under addition, $xy + x \in N$. Hence $y + 1 \in M_x$ and $M_x = N$.
(iv) Show that the well-ordering principle holds in N.

This is all fine, but where do we get the integers? Well, of course, we just tack on 0 and the negative natural numbers. Before nodding your head and shouting, "Hooray!", you must show that this new set $Z = N \cup \{0\} \cup -N$ is closed under multiplication and addition.

(v) Show that Z is closed under addition.
This is a little tricky and requires the following fact. If $m, n \in N$, then $m - n \in Z$. In particular, if $m \in N$, then $m - 1 \in Z$.
(vi) Show that Z is closed under multiplication.

So we have that Z is an ordered integral domain in which the positive elements are well-ordered.

(vii) Show that Z and the integers, $\mathbb{Z}$, are order isomorphic. That is, there exists a bijection $\phi : Z \longrightarrow \mathbb{Z}$ such that:
   (a) $\phi(x + y) = \phi(x) + \phi(y)$ for all $x, y \in Z$;
   (b) $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in Z$;
   (c) if $x < y$ in Z, then $\phi(x) < \phi(y)$ in $\mathbb{Z}$.

Appendix B

Linear Algebra

> The presentation of the material will perhaps incur the reproach of being too dogmatic in its methods. To this possible objection, we would like to make two answers. Firstly, that what the student may learn here is not designed to help him with problems he has already met but with those he will have to cope with in the future; it is therefore impossible to motivate the definitions and theorems by applications of which the reader does not know the existence as yet. Secondly, that one of the important pedagogical problems which a teacher of beginners in mathematics has to solve is to impart to his students the technique of rigorous mathematical reasoning; this is an exercise in rectitude of thought, of which it would be futile to disguise the austerity.
>
> – *Claude Chevalley, Fundamental Concepts of Algebra*

This is a book that is tilted toward analysis. However, almost all of the structures that we deal with in analysis have an underlying algebraic component, and an understanding of this algebraic component makes it a lot easier to discuss the analysis. Our approach is more user-friendly than Chevalley's. We will find that analysis is mostly about inequalities, while algebra is mostly about equalities. The fundamental algebraic ideas we discuss in this chapter concern vector spaces and linear algebra. Other algebraic structures that play a role in analysis include groups, rings, and fields. Some ideas about these were discussed in Appendix A. More can be found in the projects at the end of this chapter.

B.1. Fundamentals of Linear Algebra

The algebraic structures that are really fundamental for analysis are vector spaces (sometimes called linear spaces). Recall that a field has been defined in Definition A.6.19.

DEFINITION B.1.1. Let F be a field. A *vector space* over F is a triple $(V, +, \cdot)$ where $(V, +)$ satisfies the axioms (A1)–(A5) of Appendix A and $\cdot$ is a map from $F \times V$ to V satisfying the following properties:

(a) if $\alpha \in F$ and $\mathbf{v} \in V$, then $\alpha \cdot \mathbf{v} \in V$;
(b) if $\alpha \in F$ and $\mathbf{v}_1, \mathbf{v}_2 \in V$, then $\alpha \cdot (\mathbf{v}_1 + \mathbf{v}_2) = (\alpha \cdot \mathbf{v}_1) + (\alpha \cdot \mathbf{v}_2)$;
(c) if $\alpha, \beta \in F$ and $\mathbf{v} \in V$, then $(\alpha + \beta) \cdot \mathbf{v} = (\alpha \cdot \mathbf{v}) + (\beta \cdot \mathbf{v})$;
(d) if $\alpha, \beta \in F$ and $\mathbf{v} \in V$, then $(\alpha\beta) \cdot \mathbf{v} = \alpha \cdot (\beta \cdot \mathbf{v})$;
(e) if 1 is the multiplicative identity in F and $\mathbf{v} \in V$, then $1 \cdot \mathbf{v} = \mathbf{v}$.

The function $(\alpha, \mathbf{v}) \mapsto \alpha \cdot \mathbf{v}$ is called *scalar multiplication* and the elements of F are called *scalars*. We frequently suppress the dot $(\cdot)$.

For completeness, we restate axioms (A1)–(A5).

(A1) If $\mathbf{v}_1, \mathbf{v}_2 \in V$, then $\mathbf{v}_1 + \mathbf{v}_2 \in V$.
(A2) If $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V$, then $\mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3) = (\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3$.
(A3) If $\mathbf{v}_1, \mathbf{v}_2 \in V$, then $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$.
(A4) There exists $\mathbf{0} \in V$ such that for all $\mathbf{v} \in V$, $\mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v} = \mathbf{v}$.
(A5) For every $\mathbf{v} \in V$, there exists $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = -\mathbf{v} + \mathbf{v} = \mathbf{0}$.

Any structure satisfying (A1)–(A5) above is called an *abelian group* (see the project in Section B.6.1). Hence, a vector space is an abelian group with scalar multiplication.

EXERCISE B.1.2.

(i) If 0 is the additive identity in F and $\mathbf{0}$ is the additive identity in V, show that $0 \cdot \mathbf{v} = \mathbf{0}$ for any $\mathbf{v} \in V$. Note that this statement actually says something. It says that the additive identity in F has a property for scalar multiplication similar to multiplication by 0 in commutative rings. However, in this case, multiplication by the 0 in F gives $\mathbf{0}$ in V.
(ii) Determine whether condition (e) is necessary or follows from the other axioms.
(iii) Determine which of the statements in the definition are independent.
(iv) Do the axioms for a vector space form a completely independent system?

DEFINITION B.1.3. Let V be a vector space over a field F, and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in V$. Then a vector of the form

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_m\mathbf{v}_m$$

where $\alpha_1, \ldots, \alpha_m \in F$ is called a *linear combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

EXAMPLES B.1.4.

(1) If F is a field, $V = \{\mathbf{0}\}$ is a vector space over F.
(2) If F is a field, then F is a vector space over F with scalar multiplication being ordinary multiplication in F.
(3) Let F be a field and let $F^n = \{(x_1, x_2, \ldots, x_n) \mid x_j \in F,\ j = 1, 2, \ldots, n\}$. Addition in F^n is defined coordinatewise; that is, if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, then $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$. If

$\alpha \in F$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in F^n$, we set $\alpha \cdot \mathbf{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n)$. Then F^n is a vector space over F.

(4) Polynomial functions in one variable with real coefficients form a vector space over $\mathbb{R}$. This vector space is denoted by $\mathbb{R}[x]$. We could also consider polynomial functions in one variable with coefficients in $\mathbb{Q}$. This is a vector space over $\mathbb{Q}$ that is denoted by $\mathbb{Q}[x]$.

(5) Let X be a nonempty set, let F be a field, and let $V = \mathcal{F}(X, F)$ be the set of all functions from X to F. For $f, g \in V$, define $(f+g)(x) = f(x) + g(x)$ and, for $\alpha \in F$, define $(\alpha f)(x) = \alpha f(x)$.

(6) The real numbers $\mathbb{R}$ form a vector space over $\mathbb{Q}$.

EXERCISE B.1.5. Check that the above examples satisfy the axioms for a vector space.

REMARK B.1.6. What is a vector? That's easy to answer. A vector is an element of a vector space. Lots of people describe a vector as a quantity having magnitude and direction. This is not particularly useful in most contexts (for example, see (5) above). However, in this chapter, when we do the geometric interpretation of vectors in n-dimensional Euclidean space, it will be helpful to think of vectors this way.

EXERCISE B.1.7. Consider the collection $\mathbb{R}[x]$ of all polynomial functions with coefficients in $\mathbb{R}$. Show that $\mathbb{R}[x]$ is a vector space over $\mathbb{R}$ and also over $\mathbb{Q}$.

There are two fundamental notions that need to be discussed right at the beginning of our treatment of vector spaces. These are linear independence and linear dependence. These notions are the heart and soul of elementary vector space theory.

DEFINITION B.1.8. Let V be a vector space over a field F and let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be nonzero vectors in V. We say that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a *linearly independent* set if, for any scalars $\alpha_1, \ldots, \alpha_m \in F$ such that

$$\alpha_1 \mathbf{v}_1 + \cdots + \alpha_m \mathbf{v}_m = \mathbf{0},$$

we have $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$.

EXAMPLE B.1.9. A set containing a single nonzero vector is linearly independent.

EXERCISE B.1.10. In the vector space F^n over a field F, set $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, and, generally, $\mathbf{e}_j = (0, 0, \ldots, 1, \ldots, 0)$ where the 1 is in the j-th coordinate and 0 is in the other coordinates. Show that, for any k, $1 \leq k \leq n$, the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k\}$ is a linearly independent set.

REMARK B.1.11. An infinite set of vectors is said to be *linearly independent* if each finite subset is linearly independent.

EXERCISE B.1.12. Show that the set $\{1, x, x^2, \ldots, x^n, \ldots\}$ is a linearly independent set in $\mathbb{Q}[x]$.

DEFINITION B.1.13. Let V be a vector space over a field F, and let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be vectors in V. The set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is a *linearly dependent* set if there exist scalars $\alpha_1, \ldots, \alpha_m \in F$, not all zero, such that

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_m\mathbf{v}_m = \mathbf{0};$$

that is, a set is linearly dependent if it is not linearly independent.

EXERCISE B.1.14.

(i) Let $V = F^n$ and let $\mathbf{v}$ be any vector in $V \setminus \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Show that the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n, \mathbf{v}\}$ is a linearly dependent set in V.

(ii) Show that if $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a set of vectors in a vector space V and one of these vectors is the zero vector, then the set is a linearly dependent set.

(iii) Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be vectors in a vector space V. Show that the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly dependent set iff one of these vectors is a scalar multiple of the other.

LEMMA B.1.15. *Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a linearly dependent set of nonzero vectors in a vector space V over a field F. Then there exists k, $1 < k \leq m$, such that $\mathbf{v}_k$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}$.*

Proof. Let k be the smallest integer such that $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\}$ is a linearly independent set and $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a linearly dependent set. Then there exist scalars $\alpha_1, \ldots, \alpha_k$ such that

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$$

and $\alpha_k \neq 0$. The integer k gives the desired result.

There are two additional ideas that are closely related. One is that of a spanning set, and the second is that of a basis.

DEFINITION B.1.16. Let V be a vector space over a field F, and let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a set of vectors in V. The set S is a *spanning set* for V if every element of V can be written as a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$. A spanning set S is a *basis* for V if S is linearly independent.

REMARK B.1.17. If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a spanning set for V, we say that S *spans* V. It is possible that no finite set spans V. For example, in $\mathbb{Q}[x]$, the collection of monomials $S = \{1, x, x^2, \ldots, x^n, \ldots\}$ has the property that every element in $\mathbb{Q}[x]$ can be written as a finite linear combination of these monomials, but no proper subset of S has this property. We extend the definition of spanning set by saying that a collection S of vectors in V (possibly infinite) is a spanning set if every element of V can be written as a (finite) linear combination of elements of S. We note here that, in the absence of any convergence techniques, infinite sums of vectors do not make any sense. When we get to metric spaces, this situation will change.

EXERCISE B.1.18. Show that the vector space $\mathbb{Q}[x]$ is a countable set.

EXAMPLE B.1.19. Let F be a field and let $V = F^n$. The set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ as defined in Exercise B.1.10 is a basis for V. This is called *the standard basis*, or the *canonical basis*, for V.

What we now prove is that if a vector space has a basis with n elements, then all bases have the same number of elements. This will allow us to define the dimension of a vector space as the cardinality of a basis. The following lemma plays a crucial role.

LEMMA B.1.20 **(Exchange Lemma)**. *Suppose that V is a vector space over a field F and that $U = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is a spanning set for V. If the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a linearly independent set in V, then $n \leq m$.*

Proof. First, consider the collection $S_1 = \{\mathbf{v}_1, \mathbf{u}_1, \ldots, \mathbf{u}_m\}$. Since U spans V, we have that $\mathbf{v}_1 \in U$, or S_1 is a linearly dependent set. So, according to Lemma B.1.15, there exists an element $\mathbf{u}_{j_1} \in S_1$ that is a linear combination of $\mathbf{v}_1, \mathbf{u}_1, \ldots, \mathbf{u}_{j_1-1}$. Throw it away. Now define

$$S_2 = \{\mathbf{v}_2, \mathbf{v}_1, \mathbf{u}_1, \ldots, \mathbf{u}_{j_1-1}, \mathbf{u}_{j_1+1}, \ldots, \mathbf{u}_m\}.$$

By similar reasoning, we have that $\mathbf{v}_2 \in U$, or S_2 is linearly dependent. So, proceeding as above, we can take another $\mathbf{u}_{j_2}$ and toss it out and retain the same linear span. We can continue by putting $\mathbf{v}_3, \mathbf{v}_4, \ldots$ at the beginning of our list, and each time we do that, we can eliminate an element of U that remains on our list. We will never be able to eliminate one of the $\mathbf{v}$'s since they are linearly independent. So at each stage, we eliminate one of the $\mathbf{u}$'s and are left with a spanning set. If at some point, all the $\mathbf{u}$'s are gone and some $\mathbf{v}$'s have not been used, then a proper subset of the $\mathbf{v}$'s would be a spanning set. This contradicts the linear independence of the $\mathbf{v}$'s.

Now, we assume that our vector space V over F has a finite subset that spans. Such a vector space is called *finite-dimensional*. The next corollary proves the existence of a basis for a finite-dimensional vector space.

COROLLARY B.1.21. *Let V be a nonzero finite-dimensional vector space over a field F. Then V has a finite basis.*

Proof. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a spanning set consisting of nonzero vectors. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly dependent, then by Lemma B.1.15 there exists an integer k such that $\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\}$ is a linearly independent set and $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly dependent. Eliminating $\mathbf{v}_k$ from the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$, we still have a spanning set. Continue this process (a finite number of times). This yields a linearly independent set that spans V, that is, a basis.

Now we get to the heart of the matter.

THEOREM B.1.22. *If V is a finite-dimensional vector space over a field F, then any two bases of V have the same number of elements.*

Proof. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ are bases for V. Then by the Exchange Lemma, $n \leq m$ and $m \leq n$.

Now we can talk about an n-dimensional vector space over a field F.

DEFINITION B.1.23. Suppose V is a vector space containing a spanning set of n linearly independent vectors, $n \geq 1$. The *dimension* of V, denoted $\dim V$, is equal to n; that is, $\dim V = n$. If $V = \{\mathbf{0}\}$, we set $\dim V = 0$.

THEOREM B.1.24. *Suppose that V is an n-dimensional vector space over a field F and that $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a linearly independent set in V. Then $m \leq n$ and there exist vectors $\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n$ such that $\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$ is a basis for V.*

Proof. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis for V. If we consider the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{u}_1, \ldots, \mathbf{u}_n\}$, then, again using Lemma B.1.15 and its proof, the result follows.

DEFINITION B.1.25. If V is a vector space over a field F, then a nonempty subset $W \subseteq V$ is a *subspace* if it is closed under addition and scalar multiplication. That is, if $\mathbf{v}, \mathbf{w} \in W$, then $\mathbf{v} + \mathbf{w} \in W$, and if $\mathbf{v} \in W$ and $\alpha \in F$, then $\alpha\mathbf{v} \in W$.

THEOREM B.1.26. *Let V be a finite-dimensional vector space over a field F. Suppose that W is a subspace of V. Then $\dim W \leq \dim V$.*

Proof. Suppose that $\dim V = n$. Now consider a linearly independent set $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ in W. By Theorem B.1.24, $m \leq n$. Consider the set S of all positive integers that are the cardinalities of linearly independent sets in W. The set S is bounded above by n. Let h be the largest element of S. Let $B = \{\mathbf{w}_1, \ldots, \mathbf{w}_h\}$ be a linearly independent set in W of cardinality h. Then B must be a basis for W. Otherwise, there would be an element $\mathbf{w} \in W$ not in the span of B. Then h would not be maximal since $B \cup \{\mathbf{w}\}$ would be linearly independent. So, $h = \dim W \leq n$.

EXERCISE B.1.27.

(i) Let V be a vector space over a field F. Show that $\{\mathbf{0}\}$ and V are subspaces of V.
(ii) When is it true that the only subspaces of V are $\{\mathbf{0}\}$ and V?

EXAMPLES B.1.28.

(i) Let $V = \mathbb{Q}[x]$, and let W be the collection of all polynomials in $\mathbb{Q}[x]$ whose degree is less than or equal to a fixed nonnegative integer n. Then W is a subspace of $\mathbb{Q}[x]$.
(ii) Let $V = F^n$ and for a fixed $m \leq n$, let $W = \{v \in V \mid v = \alpha_1\mathbf{e_1} + \alpha_2\mathbf{e_2} + \cdots + \alpha_m\mathbf{e_m}, \alpha_j \in F, j = 1, \ldots, m\}$. Then W is a subspace of V.

EXERCISE B.1.29. Find the dimensions of W in the last two examples.

Most of the vector spaces which arise in analysis are not finite-dimensional and thus are called *infinite-dimensional*. We will often be dealing with vector spaces of functions whose domain and range are finite-dimensional vector spaces over $\mathbb{R}$ (or over the complex numbers $\mathbb{C}$), but the spaces of functions themselves will ordinarily be infinite-dimensional spaces. To prove the existence of a basis for an infinite-dimensional space, Zorn's lemma or some other equivalent statement is needed.

EXERCISE B.1.30. Suppose X is a set.

(i) If $V = \mathcal{F}(X, F)$, the set of functions from X to F, show that V is finite-dimensional if and only if X is finite.
(ii) Fix a subset $A \subseteq X$ and define $W(A) = \{f \in V \mid f|_A = 0\}$. Show that $W(A)$ is a subspace of V.
(iii) Can you find an infinite set X and a field F where you can write an explicit basis for $V = \mathcal{F}(X, F)$?

THEOREM B.1.31. *Let V be a nonzero vector space over a field F. Then V has a basis. That is, there exists a linearly independent subset B of V such that each element of V is a finite linear combination of elements of B.*

Proof. We apply Zorn's lemma to the collection of linearly independent sets in V. This collection can be partially ordered by inclusion. Given any totally ordered subset, the union of the elements in this subset provides a maximal element. The conclusion of Zorn's lemma gives a basis for V.

EXAMPLE B.1.32. This last theorem means that there exists a basis $B = \{\mathbf{v}_i \mid i \in I\}$ for $\mathbb{R}$ considered as a vector space over $\mathbb{Q}$. In particular, every real number can be written as a finite linear combination of elements of this basis with coefficients taken from $\mathbb{Q}$. The basis is not countable. It is called a *Hamel basis* for $\mathbb{R}$ over $\mathbb{Q}$.

B.2. Linear Transformations

One of the most important topics in the subject of linear algebra is the study of maps from one vector space over a field F to another vector space over F that preserve addition and scalar multiplication. Such maps are called *linear transformations* and they play a vital role throughout the remainder of this text.

DEFINITION B.2.1. Let V, W be vector spaces over a field F. A map $T : V \longrightarrow W$ is called a *linear transformation*, or *linear operator*, if

(a) $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$,
(b) $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$ for all $\alpha \in F$ and $\mathbf{v} \in V$.

EXERCISE B.2.2. Let V and W be vector spaces over a field F and let $T : V \longrightarrow W$ be a linear transformation. Show that $T(\mathbf{0}) = \mathbf{0}$ and $T(-\mathbf{v}) = -T(\mathbf{v})$ for all $\mathbf{v} \in V$.

EXERCISE B.2.3. Let V and W be vector spaces over a field F and let $T : V \longrightarrow W$ be a linear transformation.

(i) Show that $T(V)$ is a subspace of W.
(ii) If T is an injection, show that $T^{-1} : T(V) \longrightarrow V$ is a linear operator.

EXAMPLE B.2.4. Consider $\mathbb{R}$ as a vector space over itself and fix $a \in \mathbb{R}$. Define $T_a(x) = ax$ for $x \in \mathbb{R}$. Then T_a is a linear operator on $\mathbb{R}$. Notice that we could also consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$. Then T_a would also be a linear operator from $\mathbb{R}$ to $\mathbb{R}$, but the underlying field would be different.

EXAMPLE B.2.5. Let $V = \mathbb{R}[x]$ be considered as a vector space over $\mathbb{R}$. Define $T : \mathbb{R}[x] \longrightarrow \mathbb{R}[x]$ by $Tp(x) = p'(x)$, that is, the derivative of the polynomial p. It follows from the properties of the derivative that T is a linear transformation.

EXERCISE B.2.6. Show that $T : \mathbb{R}[x] \longrightarrow \mathbb{R}[x]$ as defined in the previous example is surjective but not injective.

A linear transformation cannot increase the dimension of a vector space. For finite-dimensional spaces, this is the content of the following theorem.

THEOREM B.2.7. *Suppose that V and W are vector spaces over a field F and that $T : V \longrightarrow W$ is a linear transformation. If V is finite-dimensional with $\dim V = n$, then $\dim T(V) \leq n$.*

Proof. It suffices to show that every subset of $T(V)$ consisting of $n+1$ elements is linearly dependent. Let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{n+1}$ be vectors in $T(V)$. Pick $\mathbf{v}_1, \ldots, \mathbf{v}_{n+1} \in V$ so that $T(\mathbf{v}_j) = \mathbf{w}_j$ for $j = 1, \ldots, n+1$. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_{n+1}\}$ is a linearly dependent set, so there exist scalars $\alpha_1, \ldots, \alpha_{n+1}$, not all zero, such that $\alpha_1\mathbf{v_1} + \cdots + \alpha_{n+1}\mathbf{v}_{n+1} = \mathbf{0}$. It follows that $T(\alpha_1\mathbf{v}_1 + \cdots + \alpha_{n+1}\mathbf{v}_{n+1}) = \alpha_1\mathbf{w}_1 + \cdots + \alpha_{n+1}\mathbf{w}_{n+1} = \mathbf{0}$. Hence the set $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{n+1}\}$ is linearly dependent.

DEFINITION B.2.8. Let V and W be vector spaces over a field F and let $T : V \longrightarrow W$ be a linear transformation. The transformation T is called a *linear isomorphism* if T is a bijection. In this case, V and W are said to be *linearly isomorphic*.

COROLLARY B.2.9. *Suppose that V and W are finite-dimensional vector spaces over a field F. If V and W are linearly isomorphic, then $\dim V = \dim W$. Moreover, if $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ is a basis for V, then $\{T(\mathbf{v_1}), \ldots, T(\mathbf{v_n})\}$ is a basis for W.*

EXERCISE B.2.10. Suppose that V and W are finite-dimensional vector spaces over a field F such that $\dim V = \dim W$. Show that V and W are linearly isomorphic.

In general, if T is a linear transformation from V to W, T is neither injective nor surjective. This leads to an important idea, the kernel of a linear transformation.

DEFINITION B.2.11. Let V and W be vector spaces over a field F, and let $T : V \longrightarrow W$ be a linear transformation. The *kernel of* T is defined by

$$\ker T = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}.$$

EXERCISE B.2.12.

(i) Show that $\ker T$ is a subspace of V.

(ii) Let $T : \mathbb{R}[x] \longrightarrow \mathbb{R}[x]$ be as in Example B.2.5, that is, $Tp(x) = p'(x)$. Find $\ker T$.

THEOREM B.2.13. *Suppose that V and W are vector spaces over a field F and that $\dim V$ is finite. Let $T : V \longrightarrow W$ be a linear transformation. Then $T(V)$ is a finite-dimensional subspace of W and $\dim V = \dim \ker T + \dim T(V)$.*

Proof. From above, we know that if $\dim V = n$, then $\dim T(V) \leq n$. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for $\ker T$. Complete this to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \ldots, \mathbf{v}_n\}$ for V by Theorem B.1.24. Then $\{T(\mathbf{v}_{k+1}), \ldots, T(\mathbf{v}_n)\}$ is a basis for $T(V)$.

DEFINITION B.2.14. Let V and W be vector spaces over a field F. If T, T_1, and T_2 are linear transformations from V to W, we define

(a) $(T_1 + T_2)(\mathbf{v}) = T_1(\mathbf{v}) + T_2(\mathbf{v})$, for $\mathbf{v} \in V$, and

(b) $(\alpha T)(\mathbf{v}) = \alpha T(\mathbf{v})$, for $\alpha \in F$.

THEOREM B.2.15. *Let V and W be vector spaces over a field F. Let $\mathcal{L}(V, W)$ denote the set of all linear transformations from V to W. Then, with the above operations, $\mathcal{L}(V, W)$ is a vector space over F.*

Proof. Clear.

EXERCISE B.2.16. Show that if $\dim V = n$ and $\dim W = m$, then $\mathcal{L}(V, W)$ is a finite-dimensional vector space with $\dim \mathcal{L}(V, W) = nm$.

The proof of this last exercise is facilitated by the use of bases in V and W. This leads to the notion of matrices representing linear transformations.

B.3. Linear Transformations and Matrices

Take finite-dimensional vector spaces V and W over a field F and $T \in \mathcal{L}(V, W)$. Suppose that $\dim V = n$, $\dim W = m$, and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ are bases for V and W, respectively. For $1 \leq k \leq n$, we can write

$$T(\mathbf{v}_k) = \sum_{j=1}^{m} a_{jk}\mathbf{w}_j,$$

where each $a_{jk} \in F$. Then, if $\mathbf{v} = \sum_{k=1}^{n} \beta_k \mathbf{v}_k$, we have

$$T(\mathbf{v}) = \sum_{k=1}^{n} \beta_k \left(\sum_{j=1}^{m} a_{jk}\mathbf{w}_j \right) = \sum_{j=1}^{m} \left(\sum_{k=1}^{n} \beta_k a_{jk} \right) \mathbf{w}_j.$$

Thus, for any $\mathbf{v} \in V$, we have written $\mathbf{v}$ with respect to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $T(\mathbf{v})$ with respect to the basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$. The coefficients in the latter expression are written in terms of the coefficients of $\mathbf{v}$ and the coefficients of $T(\mathbf{v}_k)$, $1 \leq k \leq n$. This information is encoded in a rectangular array called a *matrix*. That is, we write

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

If $\mathbf{v} \in V$ and $\mathbf{v} = \sum_{k=1}^{n} \beta_k \mathbf{v}_k$ as above, the scalars $\beta_1, \beta_2, \ldots, \beta_n$ are called the *coefficients* of $\mathbf{v}$ relative to the basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Once we have chosen a basis, it is common to represent the vector $\mathbf{v}$ by writing these coefficients in a column (or row).

Now, we compute $T(\mathbf{v})$ by computing the coefficients of $T(\mathbf{v})$ in the basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} a_{1k}\beta_k \\ \sum_{k=1}^{n} a_{2k}\beta_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}\beta_k \end{pmatrix}.$$

Observe that the matrix A has m rows and n columns (rows are horizontal and columns are vertical). This matrix is called an $m \times n$ (*m by n*) *matrix over F*. Of course, the coefficients in the matrix depend on the choice of bases in V and W. Incidentally, this might be a good time to formalize the definition of matrix.

Definition B.3.1. Let R be a commutative ring with 1. Let a_{ij} be elements of R, where $1 \leq i \leq m$ and $1 \leq j \leq n$. An $m \times n$ *matrix over R* is a rectangular array given by

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Exercise B.3.2. Let V and W be finite-dimensional vector spaces over F of dimensions n and m, respectively. Let $M_{mn}(F)$ be the collection of $m \times n$ matrices over F. We use the notation $A = (a_{ij})$ for elements of $M_{mn}(F)$. If $A = (a_{ij})$, $B = (b_{ij})$, we define $A + B = (a_{ij} + b_{ij})$ and for

$\alpha \in F$ we define $\alpha A = (\alpha a_{ij})$. Show that $M_{mn}(F)$ is a vector space over F. Find a basis for $M_{mn}(F)$. By fixing bases for V and W, give an explicit linear isomorphism between $\mathcal{L}(V, W)$ and $M_{mn}(F)$.

The above development is of particular importance in the case when $V = W$. In this case, we use the notation $\mathcal{L}(V)$ for $\mathcal{L}(V, V)$. If $\dim V = n$, then $\dim \mathcal{L}(V) = n^2$, and each element of $\mathcal{L}(V)$ can be represented by an $n \times n$ matrix relative to a chosen basis. Along with the operations of addition and scalar multiplication in $\mathcal{L}(V)$, we have composition of linear transformations. Suppose that $S, T \in \mathcal{L}(V)$. Then $S \circ T$ is defined in the usual way by $S \circ T(\mathbf{v}) = S(T(\mathbf{v}))$.

Exercise B.3.3.

(i) If $S, T \in \mathcal{L}(V)$, show that $S \circ T \in \mathcal{L}(V)$.
(ii) If $R, S, T \in \mathcal{L}(V)$, then $R \circ (S \circ T) = (R \circ S) \circ T$ (this actually follows from the associativity of composition of functions discussed in Appendix A).
(iii) If $R, S, T \in \mathcal{L}(V)$, show that $R \circ (S + T) = (R \circ S) + (R \circ T)$ and $(R + S) \circ T = (R \circ T) + (S \circ T)$.
(iv) Let $I \in \mathcal{L}(V)$ be defined by $I(\mathbf{v}) = \mathbf{v}$ for $\mathbf{v} \in V$. Show that $T \circ I = I \circ T = T$ for all $T \in \mathcal{L}(V)$.
(v) Show that if $\dim V \geq 2$, then $\mathcal{L}(V)$ is not commutative with respect to the internal law of composition $\circ$. That is, there exist $S, T \in \mathcal{L}(V)$ such that $S \circ T \neq T \circ S$.

A little vocabulary is in order here. In Appendix A, we used the terms commutative ring with 1, integral domain, and field. As pointed out there, the word "commutative" referred to the operation of multiplication. Some of the most important algebraic structures that occur in mathematics are called *algebras*.

Definition B.3.4. Let F be a field. An *algebra* over F is a set A such that A is a vector space over F and A has an internal law of composition $\circ$ satisfying the associative law and left and right distributivity. That is, for $a, b, c \in A$ and $\alpha \in F$, we have

$$\begin{aligned} a \circ (b \circ c) &= (a \circ b) \circ c, \\ a \circ (b + c) &= (a \circ b) + (a \circ c), \\ (a + b) \circ c &= (a \circ c) + (b \circ c), \\ (\alpha \cdot a) \circ b = \alpha \cdot (a \circ b) &= a \circ (\alpha \cdot b). \end{aligned}$$

An algebra A is an *algebra with identity* if there is an element $\mathbf{1} \in A$ so that $a \circ \mathbf{1} = \mathbf{1} \circ a = a$ for all $a \in A$. The algebra A is a *commutative algebra* if $a \circ b = b \circ a$ for all $a, b \in A$.

Example B.3.5. If V is a nonzero vector space over a field F, then $\mathcal{L}(V)$ is an algebra with identity which is commutative if and only if $\dim V = 1$.

EXERCISE B.3.6. Let $\mathbb{R}[x]$ be the vector space of polynomial functions in one variable over $\mathbb{R}$. Define multiplication of polynomials in the usual way. Show that $\mathbb{R}[x]$ is a commutative algebra with identity.

We have defined "multiplication" in $\mathcal{L}(V)$ as composition of linear transformations. What does this mean when we represent the elements of $\mathcal{L}(V)$ as matrices?

DEFINITION B.3.7. Suppose that V is an n-dimensional vector space over a field F and suppose $S, T \in \mathcal{L}(V)$. If $A = (a_{ij})$ and $B = (b_{ij})$ are $n \times n$ matrices over F that represent S and T relative to a given basis for V, we define matrix multiplication by

$$A \cdot B = C = (c_{ij}) \quad \text{where } c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

EXERCISE B.3.8. Show that $C = (c_{ij})$ is the matrix of the linear transformation $S \circ T$ relative to the given basis for V. Deduce that multiplication of $n \times n$ matrices over a field F is associative and that

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

is an identity for multiplication of matrices. Show further that, for $n \times n$ matrices, multiplication is left and right distributive over addition and the appropriate properties hold for scalar multiplication.

Conclusion: $M_{nn}(F)$ is an algebra with identity over F that is commutative if and only if $n = 1$.

For simplicity we will write $M_n(F)$ for $M_{nn}(F)$. Our next task is to characterize those elements of $M_n(F)$ that are *invertible*, that is, those elements that have multiplicative inverses.

B.4. Determinants

EXERCISE B.4.1. Suppose that $T \in \mathcal{L}(V)$ is a bijection. Then T has an inverse, $T^{-1} : V \longrightarrow V$. Show that $T^{-1} \in \mathcal{L}(V)$. (See Exercise B.2.3.)

Of course, $T \circ T^{-1} = T^{-1} \circ T = I$, the identity map. The problem that confronts us is the following: if $A = (a_{ij})$ is the matrix of T relative to a given basis of V, how do we find the matrix of T^{-1} relative to the same basis? We seek a matrix denoted by A^{-1}, such that $A \cdot A^{-1} = A^{-1} \cdot A = I$.

Well, this shouldn't present a great problem. All we do is write the matrix for A^{-1} as (x_{ij}). This leads to n^2 linear equations in n^2 unknowns:

$$\begin{aligned} a_{11}x_{11} + a_{12}x_{21} + \cdots + a_{1n}x_{n1} &= 1, \\ a_{21}x_{11} + a_{22}x_{21} + \cdots + a_{2n}x_{n1} &= 0, \\ &\vdots \\ a_{n1}x_{11} + a_{n2}x_{21} + \cdots + a_{nn}x_{n1} &= 0, \\ a_{11}x_{12} + a_{12}x_{22} + \cdots + a_{1n}x_{n2} &= 0, \\ &\vdots \\ a_{n1}x_{1n} + a_{n2}x_{2n} + \cdots + a_{nn}x_{nn} &= 1. \end{aligned}$$

This looks somewhat tedious, so maybe at this stage, we should just tell you the answer and consider it further in a project at the end of the chapter. But that would not be true to the nature of this book. So, we are led to the quest for determinants, one of the great discoveries in mathematics.

To begin a discussion of determinants, we must first consider the collection S_n of all self-bijections of the set $\{1, 2, \ldots, n\}$. These bijections are called *permutations of n elements*.

Example B.4.2. Let's consider two examples, namely S_2 and S_3. We can represent the elements of S_2 by arrays in which the top row lists the domain and the bottom lists the corresponding elements in the image:

$$I = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \qquad r = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Note that these arrays should be thought of not as matrices, but simply as a way to represent permutations as functions. Similarly, we can write S_3 as

$$I = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad r = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad r^2 = r \circ r = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix},$$

$$f_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \quad f_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}.$$

Exercise B.4.3.

(i) Use the basic counting principle to show that the number of elements in S_n is $n!$.

(ii) Show that the composition of two elements of S_n is also an element of S_n.

(iii) Show that the elements of S_n satisfy the associative law under composition of functions.

(iv) Define $I \in S_n$ by $I(x) = x$ for all $x \in \{1, 2, \ldots, n\}$. Show that I is an identity for S_n.

(v) If $\sigma \in S_n$, define σ^{-1} as one does for any bijection. That is, $\sigma(x) = y$ iff $\sigma^{-1}(y) = x$. Show that $\sigma \circ \sigma^{-1} = \sigma^{-1} \circ \sigma = I$.

The collection S_n with the internal law of composition $\circ$ satisfying the above properties is called a *group*. The general theory of groups is discussed in the project in Section B.6.1 at the end of this chapter. The group S_n is called *the symmetric group on n objects*.

DEFINITION B.4.4. Let σ be an element of S_n. We define the *sign* of σ by

$$\operatorname{sgn}(\sigma) = \prod_{1 \le i < j \le n} \frac{\sigma(j) - \sigma(i)}{j - i}.$$

EXERCISE B.4.5. Show that if $\sigma \in S_n$, then $sgn(\sigma) = \pm 1$.

EXERCISE B.4.6. Let σ be any map from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$. We can define $\operatorname{sgn}(\sigma)$ as above. Show that $\operatorname{sgn}(\sigma) = 0$ if and only if σ is not a bijection.

PROPOSITION B.4.7. *For $\sigma, \tau \in S_n$, we have* $\operatorname{sgn}(\sigma \circ \tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$.

Proof. We have

$$\begin{aligned} \operatorname{sgn}(\sigma \circ \tau) &= \prod_{1 \le i < j \le n} \frac{\sigma(\tau(j)) - \sigma(\tau(i))}{j - i} \\ &= \prod_{1 \le i < j \le n} \frac{\sigma(\tau(j)) - \sigma(\tau(i))}{\tau(j) - \tau(i)} \frac{\tau(j) - \tau(i)}{j - i}. \end{aligned}$$

EXERCISE B.4.8. Find the signs of the permutations in S_2 and S_3.

DEFINITION B.4.9. (a) The permutations σ in S_n for which $\operatorname{sgn}(\sigma) = 1$ are called *even permutations*. Those permutations σ for which $\operatorname{sgn}(\sigma) = -1$ are called *odd permutations*. The collection of even permutations is denoted by A_n.

(b) A permutation that interchanges two distinct elements and leaves the remaining elements fixed is called a *transposition*. The transposition which sends i to j, j to i, and leaves everything else fixed is written $(i\ j)$.

EXERCISE B.4.10.

(i) Show that a transposition is an odd permutation.

(ii) Show that every element of S_n can be decomposed as a product of transpositions.

(iii) Show that $\sigma \in S_n$ is an even permutation if and only if σ can be decomposed into an even number of transpositions. Also show that $\sigma \in S_n$ is an odd permutation if and only if σ can be decomposed into an odd number of transpositions. The number of transpositions is not unique but the parity is always the same.

(iv) Show that A_n is a group (A_n is called the *alternating group* on n objects).

(v) Show that the number of elements in A_n is $n!/2$.
(vi) Show that $\operatorname{sgn}(\sigma)$ can be defined as simply the sign of the integer $\prod_{1\le i<j\le n}(\sigma(j)-\sigma(i))$.
(vii) Show that $A_2 = \{I\}$ and that $A_3 = \{I, r, r^2\}$.
(viii) Decompose each element in S_3 as a product of transpositions.
(ix) Write explicitly as arrays the elements of A_4 and A_5.

We are now prepared to define the determinant of an $n \times n$ matrix.

DEFINITION B.4.11. Let $A = (a_{ij})$ be an $n \times n$ matrix over a field F. The *determinant of* A, denoted by $\det A$, is defined by

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}.$$

EXAMPLE B.4.12. Consider a 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Then $\det A = a_{11}a_{22} - a_{12}a_{21}$.

EXERCISE B.4.13. Write out the expression for the determinant of a 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

It should have $3! = 6$ terms.

We have two tasks ahead. The first is to illustrate the role of the determinant in computing the inverse of an $n \times n$ matrix. The second is to find some reasonable way to compute the determinant of a matrix.

DEFINITION B.4.14. Suppose $A = (a_{ij})$ is an $n \times n$ matrix over a field F. Then *the transpose of* A, denoted ${}^tA$, is the matrix obtained by reflecting A around the diagonal. That is,

$${}^tA = (a_{ji}).$$

Thus,

$$\text{if } A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \text{ then } {}^tA = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix},$$

and

$$\text{if } A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \text{ then } {}^tA = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}.$$

EXERCISE B.4.15. Let A be an $n \times n$ matrix over a field F. Show that $\det(A) = \det({}^tA)$.

EXERCISE B.4.16. Let $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. Suppose that $\det A \neq 0$. Show that A^{-1} exists, and find it.

LEMMA B.4.17. *If $A = (a_{ij})$ is an $n \times n$ matrix over a field F such that, for some m, k with $m \neq k$, the m-th row is equal to the k-th row, then* $\det A = 0$.

Proof. For any $\sigma \in S_n$ let $\tilde{\sigma} = \sigma \circ (k\ m)$, where $(k\ m)$ is the transposition defined above. Our assumption implies that

$$a_{1,\sigma(1)}a_{2,\sigma(2)} \cdots a_{n,\sigma(n)} = a_{1,\tilde{\sigma}(1)}a_{2,\tilde{\sigma}(2)} \cdots a_{n,\tilde{\sigma}(n)}.$$

On the other hand, $\operatorname{sgn}(\sigma) = -\operatorname{sgn}(\tilde{\sigma})$. This shows that $\det A = 0$.

EXERCISE B.4.18. Show that if $A = (a_{ij})$ is an $n \times n$ matrix over a field F such that, for some m, k with $m \neq k$, the m-th column is equal to the k-th column, then $\det A = 0$.

The following exercise will prove useful in our discussion of the properties of determinants.

EXERCISE B.4.19. Suppose that $A = (a_{ij})$ and $B = (b_{ij})$ are $n \times n$ matrices over a field F. Suppose further that $\sigma : \{1, 2, \ldots, n\} \longrightarrow \{1, 2, \ldots, n\}$ is not a bijection. Show that

$$\sum_{\rho \in S_n} \operatorname{sgn}(\rho) a_{1,\sigma(1)}b_{\sigma(1),\rho(1)}a_{2,\sigma(2)}b_{\sigma(2),\rho(2)} \cdots a_{n,\sigma(n)}b_{\sigma(n),\rho(n)} = 0.$$

DEFINITION B.4.20. Let $A = (a_{ij})$ be an $n \times n$ matrix over a field F. Let A_{ij} be the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and the j-th column of A. The (i, j) *cofactor* of A is the element C_{ij} of F defined by $C_{ij} = (-1)^{i+j} \det A_{ij}$.

THEOREM B.4.21. *Let $A = (a_{ij})$ be an $n \times n$ matrix over a field F. Then, for any fixed k with $1 \leq k \leq n$,*

$$\det A = a_{k1}C_{k1} + a_{k2}C_{k2} + \cdots + a_{kn}C_{kn}.$$

This is called the expansion of the determinant of A with respect to the k-th row.

EXERCISE B.4.22. Let A be a 3×3 matrix over a field F. Show that the expansion of $\det A$ with respect to any row yields the same answer you obtained in Exercise B.4.13.

Proof of the theorem. By definition, $\det A$ is the sum of products of the form

$$(*) \qquad \operatorname{sgn}(\sigma) a_{1,\sigma(1)}a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$$

where σ runs through the elements of S_n. We claim that the sum of all expressions of the form $(*)$ for which $\sigma(k) = j$ is equal to $a_{kj}C_{kj}$. If we show

this for every j with $1 \leq j \leq n$, then, summing over all j, we will get the desired result.

We have

$$\sum_{\substack{\sigma \in S_n \\ \sigma(k)=j}} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$$

$$= a_{kj} \sum_{\substack{\sigma \in S_n \\ \sigma(k)=j}} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots \widehat{a_{k,\sigma(k)}} \cdots a_{n,\sigma(n)},$$

where $\widehat{\cdot}$ indicates that the term is removed. Thus, we need to check that

$$(\#) \qquad \sum_{\substack{\sigma \in S_n \\ \sigma(k)=j}} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \cdots \widehat{a_{k,\sigma(k)}} \cdots a_{n,\sigma(n)} = (-1)^{j+k} \det(A_{kj}).$$

To compute $\det(A_{kj})$, we must first re-index the rows and columns so that the indices go from 1 to $n-1$. For this, define $\phi : \{1, 2, \ldots, n-1\} \longrightarrow \{1, 2, \ldots, \hat{k}, \ldots, n\}$ by

$$\phi(j) = \begin{cases} j & \text{for } 1 \leq j \leq k-1, \\ j+1 & \text{for } k \leq j \leq n-1. \end{cases}$$

Similarly, with k replaced by j, define a bijection $\psi : \{1, 2, \ldots, n-1\} \longrightarrow \{1, 2, \ldots, \hat{j}, \ldots, n\}$. Let $\sigma \in S_n$ be such that $\sigma(k) = j$. Then the map $\sigma \circ \phi : \{1, 2, \ldots, n-1\} \longrightarrow \{1, 2, \ldots, n\}$ does not contain j in its image. The map $\psi^{-1} \circ \sigma \circ \phi : \{1, 2, \ldots, n-1\} \longrightarrow \{1, 2, \ldots, n-1\}$ is well-defined. In fact, the map $\{\sigma \in S_n \mid \sigma(k) = j\} \longrightarrow S_{n-1}$ given by $\sigma \mapsto \psi^{-1} \circ \sigma \circ \phi$ is a bijection.

Now, recalling the definition of $\det A_{kj}$, we see that the proof of $(\#)$ follows immediately from

$$(\#\#) \qquad \operatorname{sgn}(\sigma) = (-1)^{j+k} \operatorname{sgn}(\psi^{-1} \circ \sigma \circ \phi).$$

Note that ϕ and ψ are strictly increasing maps so that $\operatorname{sgn}(\psi^{-1} \circ \sigma \circ \phi)$ coincides with the sign of $\prod_{\substack{1 \leq i < l \leq n \\ i,l \neq k}} (\sigma(l) - \sigma(i))$. Canceling on both sides of $(\#\#)$, we are left with showing that

$$\operatorname{sgn}\left(\prod_{\substack{1 \leq i < l \leq n \\ i=k \text{ or } l=k}} (\sigma(l) - \sigma(i)) \right) = (-1)^{j+k}.$$

Recalling that $\sigma(k) = j$, we see that the last product is

$$\prod_{i=1}^{k-1} (j - \sigma(i)) \cdot \prod_{l=k+1}^{n} (\sigma(l) - j) = (-1)^{k-1} \prod_{l \neq k} (\sigma(l) - j).$$

Moreover, $\operatorname{sgn}\left(\prod_{l \neq k} (\sigma(l) - j) \right)$ is clearly $(-1)^{j-1}$. Altogether, we obtain $(-1)^{k-1} \cdot (-1)^{j-1} = (-1)^{j+k}$ as desired.

EXERCISE B.4.23. Suppose that $A = (a_{ij})$ is an $n \times n$ matrix over a field F. Show that the fact that $\det A = \det {}^tA$ allows an expansion of the determinant with respect to the k-th column.

We can now assert a theorem about inverses.

THEOREM B.4.24. *If A is an $n \times n$ matrix over a field F and $\det A \neq 0$, then A has an inverse. If $\det A \neq 0$, then A^{-1} is the transpose of the cofactor matrix multiplied by the inverse of the determinant of A. That is,*

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn}. \end{pmatrix}.$$

Proof. Let $C = (C_{ij})$ and consider the product ${}^tCA$. Look at the diagonal elements in this product. The j-th diagonal element is $a_{1j}C_{1j} + \cdots + a_{nj}C_{nj} = \det A$. For the off diagonal elements, we take k, m so that $1 \leq k, m \leq n$, $k \neq m$ and consider the (m,k)-th entry of ${}^tCA$. We get $a_{1k}C_{1m} + a_{2k}C_{2m} + \cdots + a_{nk}C_{nm}$. This represents expansion of the determinant of a matrix A' that is equal to A with the exception that the m-th column has been replaced by the k-th column. By the Exercise B.4.18, this determinant is 0. Thus, if $\det A \neq 0$, then ${}^tC/\det(A)$ is the left inverse of A.

EXERCISE B.4.25. Show that ${}^tC/\det(A)$ is also a right inverse for A.

We now wish to prove that an $n \times n$ matrix A over a field F has a multiplicative inverse if and only if $\det A \neq 0$. One half of this fact was proved above. That is, if $\det A \neq 0$, then A has an inverse. The other half depends on the following important theorem.

THEOREM B.4.26. *If $A = (a_{ij})$ and $B = (b_{ij})$ are $n \times n$ matrices over a field F, then*

$$\det(AB) = \det(A) \cdot \det(B).$$

Proof. We first expand

$$\det(A) \cdot \det(B) = \sum_{\tau,\sigma \in S_n} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) a_{1,\sigma(1)} a_{2,\sigma(2)} \cdots a_{n,\sigma(n)} b_{1,\tau(1)} b_{2,\tau(2)} \cdots b_{n,\tau(n)}.$$

We re-index the product of the b_{jk}'s as follows. It is clear that $b_{1,\tau(1)} \cdots b_{n,\tau(n)} = b_{\sigma(1),\tau(\sigma(1))} \cdots b_{\sigma(n),\tau(\sigma(n))}$. For fixed σ in S_n, we see that, as τ runs through S_n, so does $\tau \circ \sigma$. Moreover, $\operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) = \operatorname{sgn}(\tau \circ \sigma)$. Hence, by letting $\rho = \tau \circ \sigma$, we have

$$\det A \cdot \det B = \sum_{\sigma \in S_n} \sum_{\rho \in S_n} \operatorname{sgn}(\rho) a_{1,\sigma(1)} b_{\sigma(1),\rho(1)} a_{2,\sigma(2)} b_{\sigma(2),\rho(2)} \cdots a_{n,\sigma(n)} b_{\sigma(n),\rho(n)}.$$

By Exercise B.4.19 this last sum will not change if we allow σ to run over all maps $\sigma : \{1, 2, \ldots, n\} \longrightarrow \{1, 2, \ldots, n\}$.

Now we let $C = AB$ and consider $\det(C)$. If $C = (c_{ij})$, then we have

$$\det(C) = \sum_{\rho \in S_n} \operatorname{sgn}(\rho) c_{1,\rho(1)} c_{2,\rho(2)} \cdots c_{n,\rho(n)}.$$

From the definition of C, we know that $c_{j,\rho(j)} = \sum_{k_j=1}^{n} a_{jk_j} b_{k_j,\rho(j)}$. This gives

$$\prod_{j=1}^{n} c_{j,\rho(j)} = \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_n=1}^{n} a_{1k_1} b_{k_1,\rho(1)} a_{2k_2} b_{k_2,\rho(2)} \cdots a_{nk_n} b_{k_n,\rho(n)}.$$

For each term $(k_1, \ldots, k_n)$ in the sum, we can define a map $\sigma\colon \{1, 2, \ldots, n\} \longrightarrow \{1, 2, \ldots, n\}$ by $1 \mapsto k_1, 2 \mapsto k_2, \ldots, n \mapsto k_n$. Notice that there are exactly n^n such maps. Hence, all maps $\sigma : \{1, 2, \ldots, n\} \longrightarrow \{1, 2, \ldots, n\}$ arise in this way. So we can index the sum by σ as we let σ run over all maps. In this way, we get

$$\det(C) = \sum_{\rho \in S_n} \sum_{\sigma} \operatorname{sgn}(\rho) a_{1,\sigma(1)} b_{\sigma(1),\rho(1))} a_{2,\sigma(2)} b_{\sigma(2),\rho(2)} \cdots a_{n,\sigma(n)} b_{\sigma(n),\rho(n)}.$$

THEOREM B.4.27. *Let A be an $n \times n$ matrix over a field F. Then A has a multiplicative inverse if and only if* $\det A \neq 0$.

Proof. Exercise.

The next exercise illustrates several of the important properties of determinants of $n \times n$ matrices.

EXERCISE B.4.28. Suppose that A is an $n \times n$ matrix over a field F.

(i) If we multiply a row or column of A by a scalar c, find the determinant of the resulting matrix.
(ii) Show that if we interchange two rows or two columns of A, then the determinant of the resulting matrix is $-\det A$.
(iii) Show that if we add a scalar multiple of any row to any other row or if we add a scalar multiple of any column to any other column, then the determinant remains unchanged.

EXERCISE B.4.29. Let A be an $n \times n$ matrix over a field F.

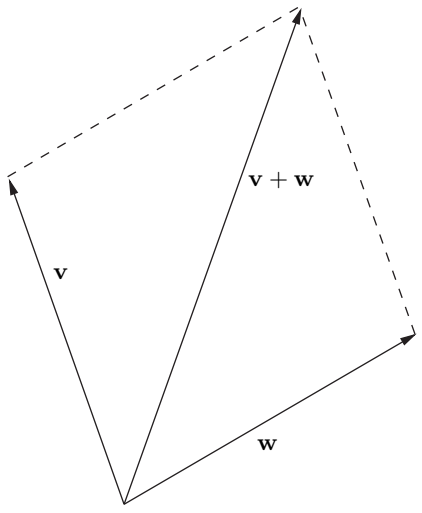
(i) If $\det A \neq 0$, show that the columns of A are linearly independent and hence form a basis for F^n.
(ii) Do the same for the rows.
(iii) If the columns of A are linearly independent, show that $\det A \neq 0$. (Hint: Consider the image of the linear transformation defined by A.)
(iv) Do the same for the rows.

B.5. Geometric Linear Algebra

We now wish to investigate the geometry of finite-dimensional vector spaces over $\mathbb{R}$. The real numbers are constructed carefully in Chapter 1. For the moment, we will use the properties of $\mathbb{R}$ that should be familiar to the reader from previous experience.

A point in $\mathbb{R}^n$ is represented by an n-tuple of elements of $\mathbb{R}$, written $\mathbf{p} = (p_1, \ldots, p_n)$, with each $p_i \in \mathbb{R}$. Geometrically, these are thought of as points in space, where the p_i's give the coordinates of the point $\mathbf{p}$. At the same time, we may consider n-tuples of real numbers as n-dimensional vectors giving the data of a direction and a magnitude, without specifying a base point from which this vector emanates. Thinking this way, we see that such vectors are elements of a vector space, $\mathbb{E}^n$, where elements can be written as $\mathbf{v} = (v_1, \ldots, v_n)$, with each $v_i \in \mathbb{R}$. We will consistently distinguish between the "points" of $\mathbb{R}^n$ and "vectors" in $\mathbb{E}^n$, since geometrically they are quite different. Observe that we are choosing the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ of Exercise B.1.10 as a basis for the vector space $\mathbb{E}^n$ and further that $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ may be written as the linear combination $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n$. Moreover, the coordinates of a point $\mathbf{p}$ are determined by a collection of mutually perpendicular coordinate axes and a distinguished point called the origin. The idea here is that vectors are free to wander around in space and points have to stay where they are.

If we want to add two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{E}^n$, we represent $\mathbf{v} + \mathbf{w}$ as the diagonal of a parallelogram pictured as



and if $\mathbf{v} = (v_1, \ldots, v_n)$ and $\mathbf{w} = (w_1, \ldots, w_n)$, then $\mathbf{v} + \mathbf{w} = (v_1 + w_1, \ldots, v_n + w_n)$. Continuing this idea, we specify that a direction $\mathbf{v}$ based at a point $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ is given by an element of $\mathbb{R}^n \times \mathbb{E}^n$. We also have a geometric operation on $\mathbb{R}^n \times \mathbb{E}^n$ which is to drag the point $\mathbf{p}$ along $\mathbf{v}$ to get a new point $\mathbf{q} = (q_1, q_2, \ldots, q_n)$. This geometric operation is algebraically

encoded in the formula

$$\mathbb{R}^n \times \mathbb{E}^n \longrightarrow \mathbb{R}^n$$
$$(\mathbf{p}, \mathbf{v}) \mapsto \mathbf{p} + \mathbf{v} = \mathbf{q}$$
$$((p_1, \ldots, p_n), (v_1, \ldots, v_n)) \mapsto (p_1 + v_1, \ldots, p_n + v_n) = (q_1, \ldots, q_n).$$

With this in mind, we have the statements

$$\text{vector} + \text{vector} = \text{vector},$$

$$\text{point} + \text{vector} = \text{point},$$

so naturally,

$$\text{point} - \text{point} = \text{vector!}$$

To make sense of this formally, given an ordered pair of points $(\mathbf{p}, \mathbf{q}) \in \mathbb{R}^n \times \mathbb{R}^n$, there is a unique vector $\mathbf{v} \in \mathbb{E}^n$ such that $\mathbf{p} + \mathbf{v} = \mathbf{q}$. This $\mathbf{v}$ represents $\mathbf{q} - \mathbf{p}$. Of course, algebraically it is given by nothing more than $\mathbf{v} = (q_1 - p_1, \ldots, q_n - p_n)$.

We are now ready to define some geometric objects in $\mathbb{R}^n$. To describe a *line*, we need a point $\mathbf{p}_0 \in \mathbb{R}^n$ and a (nonzero) direction $\mathbf{v} \in \mathbb{E}^n$.

DEFINITION B.5.1. Let $\mathbf{p}_0$ be a point in $\mathbb{R}^n$ and let $\mathbf{v}$ be a direction in $\mathbb{E}^n$. The *line ℓ through $\mathbf{p}_0$ in the direction $\mathbf{v}$* is given by

$$\ell = \{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{p} = \mathbf{p}_0 + t\mathbf{v},\, t \in \mathbb{R}\}.$$

Notice that we are using the formalism of "vector plus point equals point".

DEFINITION B.5.2. Suppose that $\mathbf{p}_0 \in \mathbb{R}^n$ and $\mathbf{v}$, $\mathbf{w}$ are linearly independent vectors in $\mathbb{E}^n$. The *plane through $\mathbf{p}_0$, spanned by $\mathbf{v}$ and $\mathbf{w}$*, is

$$\mathscr{P} = \{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{p} = \mathbf{p}_0 + t\mathbf{v} + s\mathbf{w},\, t, s \in \mathbb{R}\}.$$

More generally, we can use these ideas for other subsets of $\mathbb{R}^n$.

DEFINITION B.5.3. If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent vectors in $\mathbb{E}^n$, then we define the *k-dimensional affine subspace through $\mathbf{p}_0 \in \mathbb{R}^n$, spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_k$*, as

$$\mathbf{H} = \{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{p} = \mathbf{p}_0 + t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k, \text{ where } t_j \in \mathbb{R} \text{ and } 1 \leq j \leq k\}.$$

Note: The collection of vectors $\{t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k, t_j \in \mathbb{R}\}$ is actually a subspace of $\mathbb{E}^n$. Thus, a k-dimensional affine subspace is constructed by taking a k-dimensional subspace of $\mathbb{E}^n$ and adding it to a point of $\mathbb{R}^n$. When $k = n - 1$, $\mathbf{H}$ is called a *hyperplane* in $\mathbb{R}^n$.

DEFINITION B.5.4. If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent vectors in $\mathbb{E}^n$ with $k \leq n$ and $\mathbf{p}_0 \in \mathbb{R}^n$, we define the *$k$-dimensional parallelepiped with vertex $\mathbf{p}_0$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$* as

$$\mathbf{P} = \{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{p} = \mathbf{p}_0 + t_1\mathbf{v}_1 + \cdots + t_k\mathbf{v}_k, \text{ with } 0 \leq t_j \leq 1\}.$$

Note that if $k = n = 2$, then $\mathbf{P}$ is just a standard parallelogram in $\mathbb{R}^2$.

Much of the geometry that appears in this section will arise in more general contexts throughout mathematics. We introduce only enough here to make the reader feel comfortable in $\mathbb{R}^n$. The rich interplay between $\mathbb{R}^n$ and the vector space $\mathbb{E}^n$ is what makes life interesting.

DEFINITION B.5.5. Let V be a vector space over a field F. A *bilinear form* $\langle \cdot, \cdot \rangle$ on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \longrightarrow F$$

that satisfies linearity in both variables. That is, for all $\mathbf{v}, \mathbf{w} \in V$ and all $\alpha \in F$,

$$\begin{aligned} \langle \mathbf{v}_1 + \mathbf{v}_2, \mathbf{w} \rangle &= \langle \mathbf{v}_1, \mathbf{w} \rangle + \langle \mathbf{v}_2, \mathbf{w} \rangle, \\ \langle \alpha \mathbf{v}, \mathbf{w} \rangle &= \alpha \langle \mathbf{v}, \mathbf{w} \rangle, \\ \langle \mathbf{v}, \mathbf{w}_1 + \mathbf{w}_2 \rangle &= \langle \mathbf{v}, \mathbf{w}_1 \rangle + \langle \mathbf{v}, \mathbf{w}_2 \rangle, \\ \langle \mathbf{v}, \alpha \mathbf{w} \rangle &= \alpha \langle \mathbf{v}, \mathbf{w} \rangle. \end{aligned}$$

The form $\langle \cdot, \cdot \rangle$ is said to be *symmetric* if $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ for all $\mathbf{v}, \mathbf{w} \in V$. When $F = \mathbb{R}$, we say the form is *positive definite* if $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ for all $\mathbf{v} \in V$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$. Henceforth, we assume $F = \mathbb{R}$.

The main example of a positive definite symmetric bilinear form on $\mathbb{E}^n$ is the scalar product, or dot product.

DEFINITION B.5.6. Suppose that $\mathbf{v} = (v_1, \ldots, v_n)$ and $\mathbf{w} = (w_1, \ldots, w_n)$ are vectors in $\mathbb{E}^n$. The *scalar product* of $\mathbf{v}$ and $\mathbf{w}$ is $\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + \cdots + v_n w_n$. The scalar product is sometimes called the *dot product* and is denoted by $\mathbf{v} \cdot \mathbf{w}$. We will try our best to be consistent and use $\langle\ ,\ \rangle$.

EXERCISE B.5.7. Prove that the scalar product is a positive definite symmetric bilinear form on $\mathbb{E}^n$.

There are two concepts that arise immediately with the existence of a positive definite symmetric bilinear form. The first is the length, or norm, of a vector and the second is orthogonality.

DEFINITION B.5.8. If $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{E}^n$, then the *length*, or *norm*, of $\mathbf{v}$ is defined by

$$||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = (v_1^2 + \cdots + v_n^2)^{1/2}.$$

EXERCISE B.5.9. Prove the following properties of the norm. If $\mathbf{v}, \mathbf{w} \in \mathbb{E}^n$, then:

(i) $||\mathbf{v}|| \geq 0$;
(ii) $||\mathbf{v}|| = 0$ iff $\mathbf{v} = \mathbf{0}$;
(iii) $||\alpha \mathbf{v}|| = |\alpha|\, ||\mathbf{v}||,\ \alpha \in \mathbb{R}$;
(iv) $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$;
(v) $\|\mathbf{v} + \mathbf{w}\|^2 + \|\mathbf{v} - \mathbf{w}\|^2 = 2\left(\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2\right)$.

It is time you had the opportunity to use the quadratic formula that was so important in your earlier years.

THEOREM B.5.10 **(Cauchy-Schwarz inequality)**. *Let* $\mathbf{v}, \mathbf{w} \in \mathbb{E}^n$. *Then*

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \, \|\mathbf{w}\|.$$

Proof. Let λ be a real number. Then

$$\begin{aligned} 0 \leq \langle \mathbf{v} - \lambda \mathbf{w}, \mathbf{v} - \lambda \mathbf{w} \rangle &= \langle \mathbf{v}, \mathbf{v} \rangle - \langle \mathbf{v}, \lambda \mathbf{w} \rangle - \langle \lambda \mathbf{w}, \mathbf{v} \rangle + \langle \lambda \mathbf{w}, \lambda \mathbf{w} \rangle \\ &= ||\mathbf{v}||^2 - 2\lambda \langle \mathbf{v}, \mathbf{w} \rangle + \lambda^2 ||\mathbf{w}||^2. \end{aligned}$$

This is a quadratic polynomial in λ which is always greater than or equal to 0. It follows that, in the usual notation, $B^2 - 4AC \leq 0$. This is the desired inequality.

EXERCISE B.5.11. Prove that equality holds in the Cauchy-Schwarz inequality iff one of the vectors is a scalar multiple of the other.

This definition of the norm leads to the usual definition of Euclidean distance between two points in $\mathbb{R}^n$. Thus, if $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{R}^n$, then $d(\mathbf{p}_1, \mathbf{p}_2) = ||\mathbf{p}_1 - \mathbf{p}_2||$.

This is good enough for now. The general study of distance is carried out in Chapter 2 where we discuss metric spaces.

Since we have a positive definite symmetric bilinear form, the concept of orthogonality (or perpendicularity) in $\mathbb{E}^n$ can be formalized as follows.

DEFINITION B.5.12. Let $\mathbf{v}, \mathbf{w} \in \mathbb{E}^n$. Then $\mathbf{v}$ and $\mathbf{w}$ are said to be *orthogonal* (or *perpendicular*) if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

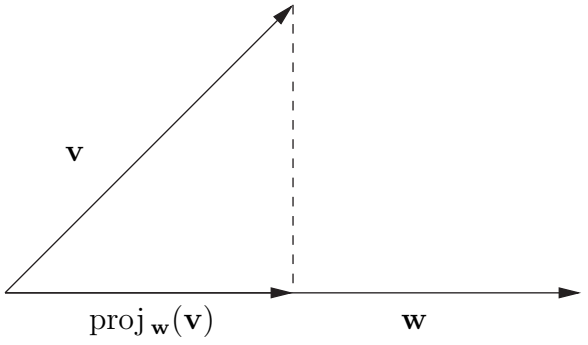
EXERCISE B.5.13.

(i) Show that the $\mathbf{0}$ vector in $\mathbb{E}^n$ is orthogonal to every vector in $\mathbb{E}^n$.
(ii) Show that the vectors in the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ are pairwise orthogonal, that is $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$ if $i \neq j$ and further that $\langle \mathbf{e}_i, \mathbf{e}_i \rangle = 1$.
(iii) If $\mathbf{v}$ is a nonzero vector in $\mathbb{E}^n$, show that the collection $W = \{\mathbf{w} \in \mathbb{E}^n \mid \langle \mathbf{w}, \mathbf{v} \rangle = 0\}$ is an $(n-1)$-dimensional subspace of $\mathbb{E}^n$.
(iv) If $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are pairwise orthogonal nonzero vectors in $\mathbb{E}^n$, show that they form a linearly independent set in $\mathbb{E}^n$.

Now, we wish to consider the angle between two nonzero vectors $\mathbf{v}, \mathbf{w}$. If the vectors are linearly dependent, that is, $\mathbf{w} = \lambda \mathbf{v}$, then the angle between them is 0 degrees if $\lambda > 0$ and 180 degrees if $\lambda < 0$. If $\mathbf{v}$ and $\mathbf{w}$ are linearly independent, we look at the plane through the origin spanned by $\mathbf{v}$ and $\mathbf{w}$. In this case, there are two angles associated to $\mathbf{v}$ and $\mathbf{w}$. One is less than 180 degrees and one is greater than 180 degrees. We take *the angle between* $\mathbf{v}$ *and* $\mathbf{w}$ to be the angle that is less than 180 degrees.

THEOREM B.5.14. *Let* $\mathbf{v}$ *and* $\mathbf{w}$ *be linearly independent vectors in* $\mathbb{E}^n$. *The angle between* $\mathbf{v}$ *and* $\mathbf{w}$ *is the unique solution* θ *to the equation*

$$\cos\theta = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{||\mathbf{v}|| \, ||\mathbf{w}||}, \quad 0 < \theta < 180^\circ.$$



Proof. This is just the law of cosines, which states that $||\mathbf{v}-\mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2 - 2||\mathbf{v}||\,||\mathbf{w}||\cos\theta$.

We next define one of the most important operations on vectors in $\mathbb{E}^n$.

DEFINITION B.5.15. Let $\mathbf{v}$ and $\mathbf{w}$ be linearly independent vectors in $\mathbb{E}^n$. The *projection of* $\mathbf{v}$ *onto* $\mathbf{w}$ is defined by

$$\text{proj}_{\mathbf{w}}(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{w}\rangle}{||\mathbf{w}||}\,\frac{\mathbf{w}}{||\mathbf{w}||}.$$

Observe that $\mathbf{w}/\|\mathbf{w}\|$ is a unit vector in the direction of $\mathbf{w}$. Hence, the projection of $\mathbf{v}$ on $\mathbf{w}$ has the same direction as $\mathbf{w}$ if $\langle \mathbf{v}, \mathbf{w}\rangle > 0$, and it has the direction of $-\mathbf{w}$ if $\langle \mathbf{v}, \mathbf{w}\rangle < 0$. Of course, if $\langle \mathbf{v}, \mathbf{w}\rangle = 0$, then $\mathbf{v}$ and $\mathbf{w}$ are orthogonal and the projection is just the zero vector. Note that the norm of $\text{proj}_{\mathbf{w}}(\mathbf{v})$ is $\|\text{proj}_{\mathbf{w}}(\mathbf{v})\| = \frac{|\langle \mathbf{v}, \mathbf{w}\rangle|}{||\mathbf{w}||}$.

EXERCISE B.5.16. Show that $\mathbf{v} - \text{proj}_{\mathbf{w}}(\mathbf{v})$ is orthogonal to $\mathbf{w}$.

Now, given a set of k linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{E}^n$, we want to construct a set $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ of linearly independent vectors such that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent and hence a basis for $\mathbb{E}^n$. In fact, we would like the vectors $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ to be orthogonal to $\mathbf{v}_1, \ldots, \mathbf{v}_k$ and, if there is more than one, to be orthogonal to each other. This can be done in a variety of ways.

First, let's look in two and three dimensions. Given a vector $\mathbf{v} = (v_1, v_2) \in \mathbb{E}^2$, $\mathbf{v} \neq \mathbf{0}$, we set $\mathbf{v}^\perp = (v_2, -v_1)$. Then, $\langle \mathbf{v}, \mathbf{v}^\perp\rangle = 0$, that is, these vectors are orthogonal. In $\mathbb{E}^3$, we have the usual notion of *cross product*. This is often defined by using determinants. For the moment, given two linearly independent vectors $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$ in $\mathbb{E}^3$, we define $\mathbf{v} \times \mathbf{w} = (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1)$.

EXERCISE B.5.17.

(i) Show that $\mathbf{v} \times \mathbf{w} \neq \mathbf{0}$.
(ii) Show that $\langle \mathbf{v}, \mathbf{v} \times \mathbf{w}\rangle = \langle \mathbf{w}, \mathbf{v} \times \mathbf{w}\rangle = \mathbf{0}$.
(iii) Show that $\mathbf{w} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{w})$.

EXERCISE B.5.18.

(i) Show that $||\mathbf{v} \times \mathbf{w}|| = ||\mathbf{v}||\,||\mathbf{w}|| \sin\theta$, where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$.

(ii) Show that $\|\mathbf{v} \times \mathbf{w}\|$ is the area of the parallelogram spanned by $\mathbf{v}$ and $\mathbf{w}$.

What we have just shown is that given a nonzero vector $\mathbf{v} \in \mathbb{E}^2$, we can find a vector $\mathbf{v}^\perp$ which is nonzero and orthogonal to $\mathbf{v}$. Obviously, the pair $\{\mathbf{v}, \mathbf{v}^\perp\}$ is a basis for $\mathbb{E}^2$. Next, given two linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{E}^3$, we constructed $\mathbf{v}_3 = \mathbf{v}_1 \times \mathbf{v}_2$ orthogonal to both of the original vectors. The set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for $\mathbb{E}^3$. This process can be generalized to $\mathbb{E}^n$.

Determinants can be used in an interesting way in the aforegoing process. For example, given a vector $\mathbf{v} = (v_1, v_2) \in \mathbb{E}^2$, we write the matrix

$$\begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 \\ v_1 & v_2 \end{pmatrix}.$$

Without being concerned about vectors in the first row and scalars in the second row, we can take the "determinant" of this matrix by expanding according to the first row and obtain $v_2\mathbf{e}_1 - v_1\mathbf{e}_2$, which is the vector $\mathbf{v}^\perp = (v_2, -v_1)$. Similarly, if we have linearly independent vectors $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$ in $\mathbb{E}^3$, we can consider

$$\begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}.$$

Taking the "determinant" of this matrix by expanding according to the first row, we get $(v_2w_3 - v_3w_2)\mathbf{e}_1 + (v_3w_1 - v_1w_3)\mathbf{e}_2 + (v_1w_2 - v_2w_1)\mathbf{e}_3$, which is $\mathbf{v} \times \mathbf{w}$.

Some of the ideas above can be extended to finite-dimensional vector spaces over any subfield of the real numbers, in particular the rational numbers. The following is a definition that is somewhat redundant but emphasizes that the ideas in the definition apply to any subfield of the real numbers.

DEFINITION B.5.19. Let F be a subfield of the real numbers, and let $V = F^n$. If $\mathbf{v} = (v_1, \ldots, v_n)$ and $\mathbf{w} = (w_1, \ldots, w_n)$ are elements of V, we say that $\mathbf{v}$ is *orthogonal* (or *perpendicular*) to $\mathbf{w}$ if $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w} = v_1w_1 + v_2w_2 + \cdots + v_nw_n = 0$. We say that $\mathbf{v}$ is a *normalized vector*, or *unit vector*, if $\langle \mathbf{v}, \mathbf{v} \rangle = v_1^2 + v_2^2 + \cdots + v_n^2 = 1$. An *orthonormal set* in V is a collection $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ of vectors in V such that $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ if $i \neq j$ and $\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1$ for all $i, j = 1, \ldots, k$. If $k = n$, the orthonormal set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is called an *orthonormal basis* for V.

THEOREM B.5.20. *Let F be a subfield of $\mathbb{R}$, and let $V = F^n$. Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}\}$ is a collection of linearly independent vectors in V,*

where $\mathbf{v}_j = (v_{j1}, v_{j2}, \ldots, v_{jn})$. *Consider the matrix*

$$\begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \\ v_{11} & v_{12} & \ldots & v_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{(n-1)1} & v_{(n-1)2} & \cdots & v_{(n-1)n} \end{pmatrix}.$$

Let $\mathbf{v}$ *be the vector obtained by taking the "determinant" of this matrix with respect to the first row. Then* $\mathbf{v}$ *is nonzero and is orthogonal to each of the vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$. *Moreover, the collection* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}, \mathbf{v}\}$ *is a basis for* V.

Proof. The vector $\mathbf{v}$ obtained by expanding with respect to the first row is simply the vector of cofactors $\mathbf{v} = (C_{11}, C_{12}, \ldots, C_{1n})$. If we replace the first row by $\mathbf{v}$, we obtain the matrix

$$A = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ v_{11} & v_{12} & \ldots & v_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{(n-1)1} & v_{(n-1)2} & \cdots & v_{(n-1)n} \end{pmatrix}.$$

We prove first that the vector $\mathbf{v}$ is orthogonal to the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$. To see this, choose i with $1 \leq i \leq n-1$. Then $\langle \mathbf{v}_i, \mathbf{v} \rangle = v_{i1}C_{11} + v_{i2}C_{12} + \cdots + v_{in}C_{1n}$. This is the determinant of the matrix obtained by replacing the first row of A by the vector $\mathbf{v}_i$. By Lemma B.4.17 this determinant is 0. Next, we establish that $\mathbf{v}$ is not the $\mathbf{0}$ vector. To do this, we note that if $C_{1j} = 0$ for all j and if we replace $\mathbf{v}$ by any vector $\mathbf{w}$ in V, then the determinant of the resulting matrix will be 0. But since $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ form a linearly independent set in V, we can extend this set to a basis in V with a vector $\mathbf{v}_n$. Replacing the vector $\mathbf{v}$ by $\mathbf{v}_n$, we get a matrix whose determinant is not 0. This is a contradiction.

EXERCISE B.5.21. If, as in the previous theorem, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}\}$ forms an orthonormal set in V, show that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}, \mathbf{v}\}$ is an orthonormal basis for V. (Hint: Consider $A^t A$ and AA^{-1}.)

EXERCISE B.5.22. Suppose that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$ are linearly independent vectors in $\mathbb{E}^n$. Take a point $\mathbf{p}_0$ in $\mathbb{R}^n$ and consider the hyperplane $\mathbf{H}$ through $\mathbf{p}_0$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$. If $\mathbf{v}$ is the vector determined in Theorem B.5.20, show that $\mathbf{H} = \{\mathbf{p} \in \mathbb{R}^n \mid \langle \mathbf{p} - \mathbf{p}_0, \mathbf{v} \rangle = 0\}$. Specialize this to obtain formulas for a line in $\mathbb{R}^2$ and for a plane in $\mathbb{R}^3$.

EXERCISE B.5.23. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$ be linearly independent vectors in $\mathbb{E}^n$. Let $\mathbf{p}_0$ be a point of $\mathbb{R}^n$. Let $\mathbf{H}$ be the hyperplane through $\mathbf{p}_0$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}$. If $\mathbf{p}$ is any point in $\mathbb{R}^n$, show that the distance from $\mathbf{p}$ to $\mathbf{H}$, that is, $\inf\{\|\mathbf{p} - \mathbf{q}\| \mid \mathbf{q} \in \mathbf{H}\}$, is given by the length of the vector $\text{proj}_{\mathbf{v}}(\mathbf{p} - \mathbf{p}_0)$ where $\mathbf{v}$ is the vector obtained in Theorem B.5.20. Specialize this to obtain formulas for the distance from a point to a line in $\mathbb{R}^2$ and from a point to a plane in $\mathbb{R}^3$.

EXERCISE B.5.24. Find a formula for the distance from a point to a line in $\mathbb{R}^n$.

Recall that our goal is to take k linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ in $\mathbb{E}^n$ for $k < n$ and find vectors $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ so that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{E}^n$. Further, we would like as much orthogonality as possible.

We know that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ can be completed to a basis for $\mathbb{E}^n$. In fact, this can be done in a systematic way by considering the collection $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ which is a spanning set. We then proceed as in the Exchange Lemma to obtain a basis for $\mathbb{E}^n$.

Of course, the $\mathbf{e}_j$'s are pairwise orthogonal, but what about the $\mathbf{v}_j$'s? This question leads to the process called *Gram-Schmidt orthogonalization*.

We begin with a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of linearly independent vectors and proceed as follows. Let $\tilde{\mathbf{v}}_1 = \mathbf{v}_1$. We continue to find vectors $\tilde{\mathbf{v}}_k$ by taking $\mathbf{v}_k$ and subtracting the projections on the vectors already constructed. More explicitly, we let

$$\begin{aligned}
\tilde{\mathbf{v}}_2 &= \mathbf{v}_2 - \operatorname{proj}_{\tilde{\mathbf{v}}_1}(\mathbf{v}_2), \\
\tilde{\mathbf{v}}_3 &= \mathbf{v}_3 - \operatorname{proj}_{\tilde{\mathbf{v}}_1}(\mathbf{v}_3) - \operatorname{proj}_{\tilde{\mathbf{v}}_2}(\mathbf{v}_3), \\
&\vdots \\
\tilde{\mathbf{v}}_k &= \mathbf{v}_k - \sum_{i=1}^{k-1} \operatorname{proj}_{\tilde{\mathbf{v}}_i}(\mathbf{v}_k).
\end{aligned}$$

It is easy to check that this set of vectors is pairwise orthogonal.

EXERCISE B.5.25. Check it, and, in addition, show that $\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \ldots, \tilde{\mathbf{v}}_k$ span the same subspace as $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$.

Now, having completed $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ to a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$, we can use the Gram-Schmidt process to obtain an orthogonal basis.

EXERCISE B.5.26. Consider the vectors $\mathbf{v}_1 = (1, 1, -1, 0)$ and $\mathbf{v}_2 = (1, 0, 0, -1)$ in $\mathbb{E}^4$, and complete this pair to an orthogonal basis as above.

The next step is to try to work with vectors whose norms are equal to one. Note that if $\mathbf{v}$ is a nonzero vector in $\mathbb{E}^n$, then $\mathbf{v}$ can be converted to a vector of norm 1, that is, a unit vector, by dividing $\mathbf{v}$ by $\|\mathbf{v}\|$.

EXAMPLE B.5.27. The collection $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is an orthonormal basis for $\mathbb{E}^n$.

EXERCISE B.5.28. Show that, given linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{E}^n$, we can transform this collection into an orthonormal set $\{\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_k\}$ that spans the same subspace. In addition, the set $\{\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_k\}$ can be completed to an orthonormal basis for $\mathbb{E}^n$.

B.6. Independent Projects

B.6.1. Groups.

In Appendix A, we introduced commutative rings with 1, integral domains, and fields. In this appendix, we have introduced vector spaces and algebras over a field, and we have defined groups in the context of the symmetric group. This project gives more details and exercises about groups.

DEFINITION B.6.1. Let G be a set with an ILC (see Definition A.6.18) $\circ : G \times G \longrightarrow G$. The pair $(G, \circ)$ is a *group* if:

(a) (associativity) for all $a, b, c \in G$, we have $(a \circ b) \circ c = a \circ (b \circ c)$;
(b) (identity) there is an *identity element* $e \in G$ such that for each $a \in G$, $e \circ a = a \circ e = a$;
(c) (solvability) for every pair $a, b \in G$, there is an element $c \in G$ such that $c \circ a = b$ and an element $d \in G$ such that $a \circ d = b$.

Here is a set of elementary exercises about groups.

EXERCISE B.6.2. Suppose that $(G, \circ)$ is a group. Prove the following statements.

(i) Show that the identity is unique.
(ii) Show that for each $a \in G$, there exists an inverse a^{-1} for a so that $a \circ a^{-1} = a^{-1} \circ a = e$.
(iii) Show that inverses are unique.
(iv) If a, b are elements of G, show that the equations $a \circ x = b$ and $x \circ a = b$ can be solved uniquely.
(v) If $a, b, c \in G$ and $a \circ b = a \circ c$, show that $b = c$.
(vi) If $a, b \in G$, show that $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$.

EXERCISE B.6.3. Show that the axioms for a group are completely independent.

EXAMPLES B.6.4. The following pairs are groups. If the ILC is addition, denoted by "+", then the identity is normally represented by the symbol "0" and the inverse of an element a is denoted by "$-a$".

Elementary examples of groups include:

(1) the ordinary integers $(\mathbb{Z}, +)$;
(2) the integers $\{1, -1\}$ under multiplication;
(3) the rational numbers $(\mathbb{Q}, +)$;
(4) the nonzero rational numbers $(\mathbb{Q}^\times, \cdot)$; and
(5) for an integer $n \geq 2$, $(\mathbb{Z}_n, +)$.
(6) The rotations of a regular n-gon about the center form a group denoted by C_n. Let r be a counterclockwise rotation through $(360/n)^\circ$, that is, $2\pi/n$ radians. The rotation group of a regular n-gon consists of I (the identity rotation), r, r^2 (the counterclockwise rotation through $(2 \cdot 360/n)^\circ$), r^3, ..., r^{n-1}. Note that $r^n = I$.

Continuing our examples of groups, we have:

(7) the group S_n of permutations of n objects;

(8) the group A_n of even permutations in S_n.
The ILC in these last two examples is composition of functions.
(9) The set of $n \times n$ matrices over a field F with addition as the ILC is a group.
(10) The set $GL_2(F) = \{A \in M_2(F) \mid \det A \neq 0\}$ is a group under multiplication of matrices. The group $GL_2(F)$ is called the 2×2 *general linear group* over F.

The first six examples of groups above fit a category of groups that are particularly easy to work with, that is, abelian, or commutative, groups.

DEFINITION B.6.5. A group $(G,\circ)$ is *abelian*, or *commutative*, if $a \circ b = b \circ a$ for all $a, b \in G$.

In the examples above, the groups S_n for $n \geq 3$, A_n for $n \geq 4$, and $GL_2(F)$ are not abelian.

REMARK B.6.6. It is often convenient to omit the ILC $\circ$ when writing a product. Thus, we can write $a \circ b = ab$. In particular, if $(G, \circ)$ is a group and k is a positive integer, we can write a^k for $a \circ a \circ \cdots \circ a$ (k-times). We can also write $a^{-k} = (a^{-1})^k$. Given all of this, we occasionally write $a \circ b$ when we feel it is useful.

The simplest groups to analyze are cyclic groups.

DEFINITION B.6.7. If G is a group and there exists an element $a \in G$ such that $G = \{a^k \mid k \in \mathbb{Z}\}$, then G is called the *cyclic group* generated by a.

Note that we do not assume, in the definition of a cyclic group, that the various powers a^k are distinct.

EXERCISE B.6.8.

(i) If G is a cyclic group of order n generated by a, show that $a^n = e$.
(ii) Show that a cyclic group of order n generated by a is also generated by a^k when k and n are relatively prime.
(iii) Show that cyclic groups are abelian and that in a cyclic group of order n, $(a^k)^{-1} = a^{n-k}$ for $1 \leq k \leq n-1$.

EXAMPLES B.6.9.

(1) The integers, $\mathbb{Z}$, under addition form an infinite cyclic group generated by 1.
(2) The even integers, $2\mathbb{Z}$, form an infinite cyclic group under addition generated by 2.
(3) Let p be a prime, and let $G = \{p^k \mid k \in \mathbb{Z}\}$. Then, G is an infinite cyclic group under multiplication generated by p.
(4) The group C_n of rotations of a regular n-gon is a cyclic group of order n generated by r (see Example B.6.4(6)).
(5) The group $(\mathbb{Z}_n, +)$ (see Definition A.6.27) is a cyclic group of order n generated by 1.

EXERCISE B.6.10. Show that the group of rational numbers under addition is not a cyclic group.

EXERCISE B.6.11. Can a cyclic group be uncountable?

DEFINITION B.6.12. Suppose that $(G, \circ)$ is a group. If the number of elements in G is finite, we write $|G| = n$, where n is the number of elements of G, and we call n the *order of* G. If G has an infinite number of elements, we say that G is an *infinite group*

DEFINITION B.6.13. Let G be a group and let $a \in G$. If there exists $m \in \mathbb{N}$ such that $a^m = e$, we say that a has *finite order in* G. If no such m exists, we say that a has *infinite order in* G. If a is of finite order, then $n = \min\{m \in \mathbb{N} \mid a^m = e\}$ is called the *order of* a *in* G.

Note that the order of an element a is 1 if and only if a is the identity element.

EXERCISE B.6.14.

(i) If G is a finite group, show that the order of an element in G is less than or equal to the order of G.
(ii) Find the orders of the elements in C_{12} and S_4.
(iii) Does there exist a group with elements of both finite and infinite order?

We next need the notion of a subgroup.

DEFINITION B.6.15. Let $(G, \circ)$ be a group. A nonempty subset H of G is a *subgroup* if the pair $(H, \circ)$ is a group. If H is a subgroup of G, with $H \neq G$ and $H \neq \{e\}$, we say that H is a *proper subgroup of* G.

REMARK B.6.16. Thus, we require that $\circ$ is an ILC on H, that the identity of the group e is an element of H, and that if an element a is in H, then a^{-1} is in H. Observe that associativity is automatic for the elements of H because associativity holds for the elements of G. In this situation, we say that associativity is *inherited* from G.

EXERCISE B.6.17. Show that a subgroup of a cyclic group is a cyclic group. That is, if G is the cyclic group generated by a and H is a subgroup of G, then there exists an element $k \in \mathbb{N}$ such that a^k generates H. This means that for every $h \in H$, there exists $j \in \mathbb{Z}$ such that $(a^k)^j = h$.

EXERCISE B.6.18. Show that every element a generates a cyclic subgroup.

EXAMPLES B.6.19.

(1) The groups $\{e\}$ and G are subgroups of G.
(2) The group $(2\mathbb{Z}, +)$, the even integers, is a subgroup of $(\mathbb{Z}, +)$, the additive group of integers.
(3) The group $(n\mathbb{Z}, +)$, the integer multiples of n, is a subgroup of $(\mathbb{Z}, +)$.
(4) The group $\{I, r^3, r^6\}$ is a subgroup of C_9, the group of rotations of a regular 9-gon. Note that this is an example of Exercise B.6.17.

(5) The groups $\{I, (12)\}$, $\{I, (23)\}$, and $\{I, (13)\}$ are subgroups of S_3.
(6) The set $SL_2(\mathbb{Q}) = \{A \in GL_2(\mathbb{Q}) \mid \det(A) = 1\}$ is a subgroup of $GL_2(\mathbb{Q})$. This is called the 2×2 *special linear group* over $\mathbb{Q}$.
(7) The group A_n is a subgroup of S_n.

EXERCISE B.6.20.

(i) Suppose that G is a group and H is a nonempty subset of G. Show that H is a subgroup of G iff for every $a, b \in H$, $a^{-1}b \in H$.
(ii) Suppose that G is a finite group and H is a nonempty subset of G. Show that H is a subgroup of G iff H is closed under multiplication.

DEFINITION B.6.21. Suppose that G_1 and G_2 are groups. A map $\phi : G_1 \longrightarrow G_2$ is a *homomorphism* if $\phi(ab) = \phi(a)\phi(b)$ for all $a, b \in G_1$. That is, a homomorphism preserves multiplication. In general, a homomorphism of algebraic structures is a map that preserves all the operations of that structure. For example, a homomorphism of rings (or fields) preserves addition and multiplication.

If a homomorphism ϕ is a surjection, then ϕ is called an *epimorphism.* If a homomorphism ϕ is an injection, then ϕ is called a *monomorphism.* If a homomorphism ϕ is a bijection, then ϕ is called an *isomorphism.*

Of course, group homomorphisms have many of the properties of linear transformations.

PROPOSITION B.6.22. *Let G_1 and G_2 be groups and let $\phi : G_1 \longrightarrow G_2$ be a homomorphism.*

(i) *If e_1 is the identity in G_1, then $\phi(e_1) = e_2$, the identity in G_2.*
(ii) *If $a \in G_1$, then $\phi(a^{-1}) = (\phi(a))^{-1}$.*

Proof. You do it.

EXAMPLE B.6.23. Let $G_1 = G_2 = (\mathbb{Z}, +)$. For $n \in \mathbb{N}$ with $n \geq 2$, define $\phi_n(a) = na$. Then ϕ_n is a homomorphism, and in fact ϕ_n is a monomorphism. If we let $G_2 = (n\mathbb{Z}, +)$, then ϕ_n is an isomorphism.

EXERCISE B.6.24.

(i) Let G be a finite cyclic group of order n generated by a. Define $\phi : G \longrightarrow G$ by $\phi(a^j) = a^{2j}$. Show that ϕ is a homomorphism. Determine those values of n for which ϕ is a monomorphism, epimorphism, or homomorphism.
(ii) Let n be a natural number with $n \geq 2$. Define $\phi : \mathbb{Z} \longrightarrow \mathbb{Z}_n$ by $\phi(k) = k \mod n$. Show that ϕ is an epimorphism.

For groups, we have a situation that is analogous to the kernel of a linear transformation in vector spaces.

DEFINITION B.6.25. Let G_1, G_2 be groups and let $\phi : G_1 \longrightarrow G_2$ be a homomorphism. The *kernel of ϕ* is the subset of G defined by

$$\ker \phi = \{x \in G_1 \mid \phi(x) = e_2, \text{ the identity in } G_2\}.$$

EXERCISE B.6.26.

(i) Show that $\ker \phi$ is a subgroup of G_1.
(ii) Show that $\phi(G_1)$ is a subgroup of G_2.
(iii) Find $\ker \phi$ in the examples and exercises for homomorphisms above.

The kernel of a homomorphism $\phi : G_1 \longrightarrow G_2$ is a subgroup with special properties. In particular, if $a, b \in G_1$ and $b \in \ker \phi$, then $aba^{-1} \in \ker \phi$.

DEFINITION B.6.27. Let G be a group and let H be a subgroup of G. The subgroup H is called a *normal subgroup of* G if, for each $a \in G$, we have $aHa^{-1} = \{aba^{-1} \mid b \in H\} = H$.

EXERCISE B.6.28.

(i) Let G_1, G_2 be groups and let $\phi : G_1 \longrightarrow G_2$ be a homomorphism. Show that $\ker \phi$ is a normal subgroup of G_1.
(ii) Show that any subgroup of an abelian group is a normal subgroup.
(iii) Show that A_n is a normal subgroup of S_n.
(iv) Show that if G is a finite group and H is a subgroup of G, then H is a normal subgroup of G if and only if $aHa^{-1} \subseteq H$.

Take the collection $M_n(F)$ of $n \times n$ matrices over a field F, and consider

$$GL_n(F) = \{x \in M_n(F) \mid \det(x) \neq 0\}.$$

EXERCISE B.6.29. Prove that $GL_n(F)$ is a group with the ILC given by multiplication of matrices. The group $GL_n(F)$ is called the $n \times n$ *general linear group over* F.

Here is a list of some interesting subgroups of $GL_n(F)$.

(1) The $n \times n$ *special linear group over* F,

$$SL_n(F) = \{x \in GL_n(F) \mid \det(x) = 1\}.$$

EXERCISE B.6.30. Show that $SL_n(F)$ is a normal subgroup of $GL_n(F)$.

(2) The *upper triangular* matrices in $GL_n(F)$,

$$B = \{(b_{ij}) \in GL_n(F) \mid b_{ij} = 0 \text{ if } i > j\}.$$

(3) The *upper triangular unipotent* matrices in $GL_n(F)$,

$$N = \{(b_{ij}) \in B \mid b_{jj} = 1, \ \ j = 1, \ldots, n\}.$$

(4) The $n \times n$ *diagonal* matrices in $GL_n(F)$,

$$A = \{(a_{ij}) \in GL_n(F) \mid a_{ij} = 0 \text{ if } i \neq j\}.$$

(5) The $n \times n$ *orthogonal group over* F,

$$O_n(F) = \{x \in GL_n(F) \mid x^t x = I, \text{ the } n \times n \text{ identity matrix}\}.$$

B.6.2. Direct Products, Semidirect Products, and Subgroups of the Orthogonal Group.

We now define a class of subgroups of $O_n(F)$. Let $G_n = \{x \in O_n(F) \mid$ there is exactly one 1 in each row and column and zeroes elsewhere$\}$.

EXERCISE B.6.31. Show that G_n is actually a subgroup of $O_n(F)$.

For example,

$$G_3 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \right\}.$$

EXERCISE B.6.32.

(i) Show that G_n is isomorphic to S_n.

(ii) If $g \in G_n$ and g corresponds to the permutation $\sigma \in S_n$, then the action of g as a permutation can be written as follows. If $a = (a_1, a_2, \ldots, a_n)$ and

$$A = \begin{pmatrix} a_1 & 0 & \ldots & 0 \\ 0 & a_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_n \end{pmatrix},$$

show that

$$gAg^{-1} = \begin{pmatrix} a_{\sigma(1)} & 0 & \ldots & 0 \\ 0 & a_{\sigma(2)} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{\sigma(n)} \end{pmatrix}.$$

The next few definitions involve creating new groups from old groups.

DEFINITION B.6.33. Let G_1 and G_2 be groups and consider the Cartesian product $G_1 \times G_2$. Define an ILC on $G_1 \times G_2$ as follows. If $(x_1, y_1), (x_2, y_2) \in G_1 \times G_2$, then $(x_1, y_1) \circ (x_2, y_2) = (x_1 \circ x_2, y_1 \circ y_2)$.

EXERCISE B.6.34.

(i) Show that $(G_1 \times G_2, \circ)$ is a group.

(ii) Show that $G_1 \times G_2$ is commutative if and only if both G_1 and G_2 are commutative.

(iii) Show that $G_1 \times \{e_2\}$ and $\{e_1\} \times G_2$ are subgroups of $G_1 \times G_2$ that are isomorphic to G_1 and G_2, respectively.

(iv) Show that $G_1 \times \{e_2\}$ and $\{e_1\} \times G_2$ are normal subgroups of $G_1 \times G_2$.

(v) If H_1 and H_2 are subgroups of G_1 and G_2, respectively, show that $H_1 \times H_2$ is a subgroup of $G_1 \times G_2$.

(vi) Does every subgroup of $G_1 \times G_2$ have the form given in (v)?

DEFINITION B.6.35. The group $(G_1 \times G_2, \circ)$ is called the *direct product* of G_1 and G_2.

It is clear that the idea of direct product can be extended to a finite number of groups.

EXAMPLES B.6.36.

(i) The group $(\mathbb{R}^n, +)$ is the direct product of $(\mathbb{R}, +)$ with itself n times.
(ii) The multiplicative group $\mathbb{R}^\times$ of nonzero real numbers has two subgroups of direct interest, namely $\mathbb{R}^\times_+$, the multiplicative group of positive real numbers, and the group $\{1, -1\}$. Clearly, $\mathbb{R}^\times = \mathbb{R}^\times_+ \times \{1, -1\}$.

DEFINITION B.6.37. Let G be a group and let H and K be subgroups of G, with H a normal subgroup. If $H \cap K = \{e\}$ and $HK = \{h \circ k \mid h \in H, k \in K\} = G$, then G is called the *semidirect product* of H and K, and we write $G = H \rtimes K$.

Observe that multiplication in the semidirect product is carried out as follows. We drop the circle for the ILC for convenience. If $h_1, h_2 \in H$ and $k_1, k_2 \in K$, then $(h_1k_1)(h_2k_2) = (h_1k_1h_2k_1^{-1})(k_1k_2)$ with $h_1k_1h_2k_1^{-1} \in H$ and $k_1k_2 \in K$.

EXAMPLES B.6.38. We need one more example of groups illustrating the semidirect product. Let P_n be the set of $n \times n$ matrices that have exactly one 1 or one -1 in each row and column and zeroes elsewhere. If $H = G_n$, considered as a subgroup of $O_n(F)$, and $\Delta_n = \{(a_{jk}) \mid a_{jj} = \pm 1$ for $1 \leq j \leq n$ and $a_{jk} = 0$ for $j \neq k\}$, then $P_n = \Delta_n \rtimes G_n$.

EXERCISE B.6.39. Show that P_n is a subgroup of $O_n(F)$ and that G_n and Δ_n satisfy the conditions for a semidirect product as subgroups of P_n.

EXAMPLE B.6.40. One of the standard examples of a semidirect product is the group of symmetries of a regular n-gon in the plane. The symmetries come in two flavors. First, there are the rotations of the n-gon, which form a cyclic group C_n of order n. Second, there are the reflections about a line of symmetry in the n-gon. The lines of symmetries for reflections are different for n even or odd. If n is odd, the lines of symmetry (or reflection) are those which join a vertex to the midpoint of the opposite side. If n is even, the lines of symmetry either connect two opposite vertices or form the bisectors of two opposite sides. The group of symmetries of a regular n-gon is made up of the n rotations in C_n and the n reflections just indicated. The symmetry group of a regular n-gon is called the *dihedral group of order* $2n$ and is denoted D_{2n}.

EXERCISE B.6.41. Let K be the subgroup of D_{2n} generated by a single reflection. Show that $D_{2n} = C_n \rtimes K$.

EXERCISE B.6.42. Show that D_{2n} is a subgroup of $O_2(\mathbb{R})$.

EXERCISE B.6.43.

(i) Let G be a finite group of order $2m$ and let H be a subgroup of G of order m. Show that H is a normal subgroup of G and that if m is odd, then G is a semidirect product of H with a subgroup of order 2.

(ii) Show that there is a subgroup K of S_n so that $S_n = A_n \rtimes K$.

EXERCISE B.6.44. Show that in the examples following Exercise B.6.29, $B = N \rtimes A$.

B.6.3. Orthogonal Transformations in Euclidean Space.

The orthogonal transformations in $\mathbb{R}^n$ play an important role in analysis.

DEFINITION B.6.45. An *orthogonal transformation* on $\mathbb{R}^n$ is a linear transformation $T : \mathbb{E}^n \longrightarrow \mathbb{E}^n$ with the property that $\langle T\mathbf{v}, T\mathbf{w}\rangle = \langle \mathbf{v}, \mathbf{w}\rangle$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{E}^n$. We say that T preserves the bilinear form $\langle \cdot\,, \cdot \rangle$.

EXERCISE B.6.46. Show that an orthogonal linear transformation is distance-preserving. That is, for any $\mathbf{v}, \mathbf{w} \in \mathbb{E}^n$, $\|T\mathbf{v} - T\mathbf{w}\| = \|\mathbf{v} - \mathbf{w}\|$. In particular, for any $\mathbf{v} \in \mathbb{E}^n$, we have $\|T\mathbf{v}\| = \|\mathbf{v}\|$.

EXERCISE B.6.47. Let M be the matrix of a linear transformation on $\mathbb{E}^n$ relative to the standard basis. Show that, for any pair of vectors $\mathbf{v}, \mathbf{w} \in \mathbb{E}^n$, we have $\langle M\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, {}^tM\mathbf{w}\rangle$. Show in particular that if M is the matrix of an orthogonal linear transformation relative to the standard basis, then $M\,{}^tM = {}^tMM = I$. That is, $M \in O_n(\mathbb{R})$.

EXAMPLE B.6.48. Let

$$M = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

be the matrix of an orthogonal linear transformation on $\mathbb{E}^2$ relative to the standard basis, that is, $M \in O_2(\mathbb{R})$. From the exercise above, it follows that $a_{11}^2 + a_{21}^2 = a_{12}^2 + a_{22}^2 = a_{11}^2 + a_{12}^2 = a_{21}^2 + a_{22}^2 = 1$. Also, $a_{11}a_{12} + a_{21}a_{22} = a_{11}a_{21} + a_{12}a_{22} = 0$. It is now immediate that there is some θ with $0 \le \theta < 2\pi$ such that

$$M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

or

$$M = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

PROPOSITION B.6.49. *If $M \in O_n(\mathbb{R})$, then* $\det M = \pm 1$.

Proof. This follows from the fact that $\det {}^tM = \det M$, so that if $M \in O_n(\mathbb{R})$, then $(\det M)^2 = \det M \det {}^tM = \det M\,{}^tM = \det I = 1$.

The collection $SO_n(\mathbb{R}) = \{M \in O_n(\mathbb{R}) \mid \det M = 1\}$ is a subgroup of $O_n(\mathbb{R})$ called the *special orthogonal group*. The elements of $SO_n(\mathbb{R})$ are called generalized rotations.

We can write

$$O_n(\mathbb{R}) = SO_n(\mathbb{R}) \cup \begin{pmatrix} -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} SO_n(\mathbb{R}).$$

The elements of the set

$$\begin{pmatrix} -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} SO_n(\mathbb{R})$$

are called generalized reflections and have determinant equal to -1.

Exercise B.6.50.

(i) If $M \in O_2(\mathbb{R})$ and $M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, show that M is a counterclockwise rotation around the origin through an angle θ. Obviously, $\det M = 1$.

(ii) If $M \in O_2(\mathbb{R})$ and $M = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, show that M is a reflection across a line through the origin. Here, $\det M = -1$. In particular, determine the angle that the line of reflection makes with the positive x-axis.

(iii) Finally, show that $SO_2(\mathbb{R}) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \;\middle|\; 0 \le \theta < 2\pi \right\}$ and that $O_2(\mathbb{R})$ is the union of the matrices from parts (i) and (ii).

(iv) Write the dihedral group D_{2n} as a subgroup of matrices in $O_2(\mathbb{R})$.

It is possible to write some "rotations" and "reflections" in $\mathbb{R}^n$ very explicitly. For example, in $O_4(\mathbb{R})$ the matrix

$$M = \begin{pmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & \cos\phi & -\sin\phi \\ 0 & 0 & \sin\phi & \cos\phi \end{pmatrix}$$

is the composition of a rotation around the x_1-x_2-plane through the angle ϕ with a rotation around the x_3-x_4-plane through the angle θ. This is an example of a generalized rotation.

As an example of a generalized reflection, if we have $n-1$ pairwise orthogonal nonzero vectors in $\mathbb{R}^n$, we can produce an n-th nonzero vector v, orthogonal to each of the original $n-1$ vectors, by Theorem B.5.20. In this

case reflection through the hyperplane spanned by the original set of $n-1$ orthogonal vectors is given by $T_v(w) = w - 2\frac{\langle v,w\rangle}{\langle v,v\rangle}w$.

Exercise B.6.51.

(i) Show that T_v is an orthogonal transformation that is a generalized reflection on $\mathbb{R}^n$.
(ii) Find the matrix of T_v relative to the standard basis.

When $n = 3$, we can be quite precise about the nature of the elements of $O_3(\mathbb{R})$. If T is an element of $SO_3(\mathbb{R})$, then we can find a line through the origin so that T is a rotation around that line. It is then clear that any reflection in $O_3(\mathbb{R})$ can be written as a rotation around a line in $\mathbb{R}^3$ through the origin combined with reflection through the origin, that is, multiplication by $-I$. From this we conclude $O_3(\mathbb{R}) = SO_3(\mathbb{R}) \times \{I, -I\}$.

We finish this project with the symmetry groups of the regular polyhedra in $\mathbb{R}^3$. The regular polyhedra are the regular tetrahedron, the cube, the regular octahedron, the regular dodecahedron, and the regular icosahedron. Since the octahedron is dual to the cube and the icosahedron is dual to the dodecahedron, we need only work with the tetrahedron, the cube, and the dodecahedron. In each case, we can obtain an upper bound on the number of symmetries by proceeding as follows. Each vertex must be mapped to a vertex and the images of the sides adjacent to a vertex must be adjacent to the image of that vertex, although they can be permuted after the symmetry map is applied. For a tetrahedron, this gives an upper bound of $4 \times 6 = 24$ possible symmetries. In this case, the symmetries are in one-to-one correspondence with the permutations of the vertices, and the symmetry group is S_4.

Exercise B.6.52.

(i) Write the twenty-four orthogonal matrices that represent the symmetries of the tetrahedron relative to the standard basis.
(ii) Show that the rotations in the symmetry group of the tetrahedron form the group A_4.
(iii) Which of the reflections in the symmetry group of a regular tetrahedron can be realized as reflections through a plane?

Now consider the symmetry group of the cube with vertices at the eight points $(\pm 1, \pm 1, \pm 1)$. The rotations of this cube can be realized as rotations around the x-axis, rotations around the y-axis, rotations around the z-axis, rotations around the diagonals of the cube, and finally rotations around the lines through the origin and the midpoints of opposite edges.

Exercise B.6.53.

(i) Show that rotations around a coordinate axis have order 2 or 4, rotations around a diagonal have order 3, and rotations around a line through the origin connecting the midpoints of opposite edges have order 2.

(ii) Write the matrices of the twenty-four rotational symmetries of the cube.

For example, the rotations around the z-axis are I, $R = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $R^2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $R^3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

If we rotate around the diagonal adjoining $(1,1,1)$ and $(-1,-1,-1)$, we obtain the nonidentity matrices $R = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and $R^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$.

If we rotate around the line connecting $(1,0,1)$ and $(-1,0,-1)$, then we obtain the matrix $R = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

We obtain the twenty-four reflections in the symmetry group of a cube by multiplying the twenty-four rotations by $-I$.

We conclude from the above discussion that the symmetry group of a cube is isomorphic to $S_4 \times \mathbb{Z}_2$.

Exercise B.6.54.

(i) Show that the group of rotations of a dodecahedron is isomorphic to A_5 and the symmetry group of a regular dodecahedron is isomorphic to $A_5 \times \mathbb{Z}_2$.

(ii) Write the matrices for the sixty rotational symmetries of a regular dodecahedron.

Bibliography

[1] Lars V. Ahlfors. *Complex analysis*. McGraw-Hill Book Co., New York, third edition, 1978. An introduction to the theory of analytic functions of one complex variable, International Series in Pure and Applied Mathematics.

[2] T. Apostol. *Mathematical analysis*. Addison-Wesley Publishing Co., Reading, Massachusetts-London-Don Mills, Ontario, second edition, 1974.

[3] J. Dieudonné. *Foundations of modern analysis*. Pure and Applied Mathematics, Vol. X, Academic Press, New York-London, 1960.

[4] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[5] Gerald B. Folland. *Advanced Calculus*. Prentice Hall, Upper Saddle River, New Jersey, 2001.

[6] Fernando Q. Gouvêa. *p-adic numbers*. Universitext. Springer-Verlag, Berlin, second edition, 1997. An introduction.

[7] Paul R. Halmos. *Measure theory*. D. Van Nostrand Company, Inc., New York, New York, 1950.

[8] Thomas Hawkins. *Lebesgue's theory of integration. Its origins and development*. The University of Wisconsin Press, Madison, Wisconsin-London, 1970.

[9] Edwin Hewitt and Kenneth A. Ross. *Abstract harmonic analysis. Vol. I: Structure of topological groups. Integration theory, group representations*. Die Grundlehren der mathematischen Wissenschaften, Band 115. Academic Press Inc., Publishers, New York, 1963.

[10] Edwin Hewitt and Kenneth A. Ross. *Abstract harmonic analysis. Vol. II: Structure and analysis for compact groups. Analysis on locally compact Abelian groups*. Die Grundlehren der mathematischen Wissenschaften, Band 152. Springer-Verlag, New York, 1970.

[11] Yitzhak Katznelson. *An introduction to harmonic analysis*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2004.

[12] A. N. Kolmogorov and S. V. Fomin. *Introductory real analysis*. Revised English edition. Translated from the Russian and edited by Richard A. Silverman. Prentice-Hall Inc., Englewood Cliffs, New York, 1970.

[13] Serge Lang. *Undergraduate analysis*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1997.

[14] Serge Lang. *Algebra*. Volume 211 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2002.

[15] Joram Lindenstrauss and Lior Tzafriri. *Classical Banach spaces. I*. Springer-Verlag, Berlin, 1977. Sequence spaces, Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 92.

[16] Joram Lindenstrauss and Lior Tzafriri. *Classical Banach spaces. II.* Volume 97 of *Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas]*. Springer-Verlag, Berlin, 1979. Function spaces.

[17] Lynn H. Loomis. *An introduction to abstract harmonic analysis*. D. Van Nostrand Company, Inc., Toronto-New York-London, 1953.

[18] Lynn H. Loomis and Shlomo Sternberg. *Advanced calculus*. Jones and Bartlett Publishers, Boston, Massachusetts, revised edition, 1990.

[19] William Fogg Osgood. *Functions of real variables; Functions of a complex variable*. Bound in one volume. Chelsea Publishing Co., New York, New York, 1958.

[20] Mark A. Pinsky. *Introduction to Fourier analysis and wavelets*. Volume 102 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 2009. Reprint of the 2002 original.

[21] L. S. Pontryagin. *Topological groups*. Translated from the second Russian edition by Arlen Brown. Gordon and Breach Science Publishers, Inc., New York, 1966.

[22] Charles Chapman Pugh. *Real mathematical analysis*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 2002.

[23] H. L. Royden. *Real analysis*. Macmillan Publishing Company, New York, third edition, 1988.

[24] Walter Rudin. *Principles of mathematical analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, third edition, 1976.

[25] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[26] Walter Rudin. *Fourier analysis on groups*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1990. Reprint of the 1962 original, A Wiley-Interscience Publication.

[27] Paul J. Sally, Jr. *Tools of the trade*. Introduction to advanced mathematics. American Mathematical Society, Providence, Rhode Island, 2008.

[28] J.-P. Serre. *A course in arithmetic*. Springer-Verlag, New York, 1973. Translated from the French, Graduate Texts in Mathematics, No. 7.

[29] Lawrence C. Washington. *Introduction to cyclotomic fields*. Volume 83 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1997.

Index of Terminology

C^∞ function, 147, 162
C^k function, 147, 162
T_0 topological space, 67
T_1 topological space, 67
T_2 topological space, 67
T_3 topological space, 67
T_4 topological space, 67
σ-algebra, 192–199, 205, 209, 223
σ-algebra generated by a collection of sets, 193, 223, 260
σ-finite measure, 200
σ-finite measure space, 194, 223, 224
$m \times n$ matrix, 322
n-sphere, 40
n-th root of unity, 23
p-adic absolute value, 85
p-adic completion of $\mathbb{Q}$, 86
p-adic expansion, 90, 101, 247
p-adic field, 86, 94, 101, 194, 259
p-adic numbers, 94

abelian group, 341
absolute continuity of the integral, 213, 215, 216
absolute convergence, 241, 244
absolute value, 9, 13, 19, 74, 89, 94, 101
absolutely continuous, 246
absolutely continuous function, 51, 52, 202
absolutely convergent series, 27
accumulation point, 15, 22, 40, 65
adjoint operator, 116, 120, 121
affine subspace, 333
algebra, 323
algebra with identity, 323
algebraic closure, 71, 72, 76, 101
algebraic extension, 19, 71
algebraic number, 24, 72
algebraic of degree n, 148, 149
algebraically closed, 24, 72
almost everywhere, 182, 207
analytic function, 158, 255
angle between vectors, 335
antipodal points, 77
Archimedean ordered field, 4, 25
Archimedean property, 4
arithmetic-geometric mean inequality, 173
Arzela-Ascoli theorem, 93
associative, 293
associativity of composition of functions, 293
automorphism group of a field, 18, 33, 101
automorphism of a field, 18, 94, 101
axiom, 310
axiom system, 310

Baire Category Theorem, 78, 81, 109
Banach algebra, 108
Banach space, 79, 103, 219
Banach, Stefan, 101
Banach-Steinhaus theorem, 110
base for a topological space, 66, 68, 198
basic additive character, 259, 260
basis, 121
basis of a vector space, 104, 155, 316
Bessel's inequality, 132, 240, 244, 245
Big Hahn-Banach Theorem, 112–114
bijective function, 292
bilinear form, 334
binary operations, 288
binomial coefficient, 284
binomial theorem, 284
Bolzano-Weierstrass lemma, 14
Bolzano-Weierstrass theorem, 15, 22, 46
Borel measure, 199, 200, 205, 214, 260
Borel set, 198–201, 227, 238
Borsuk number, 77
boundary, 40, 42, 65, 187
boundary point, 40
bounded, 72
bounded above, 2, 72
bounded below, 2, 72

Bounded Inverse Theorem, 110–112
bounded linear functional, 108, 117
bounded linear operator, 79, 106, 237, 269
bounded sequence, 10, 14, 21
bounded set, 2, 21, 48, 57
bounded variation, 52
Bourbaki, Nicolas, 33

canonical basis, 317
Cantor function, 53, 202
Cantor set, 16, 53, 80, 185, 202, 206
cap, 55
cardinal number of a finite set, 298
Cartesian product, 278
Cartesian product of groups, 346
Cartesian product of sets, 277
Cauchy criterion, 14, 21
Cauchy sequence, 10, 14, 21, 44
Cauchy's Mean Value Theorem, 145
Cauchy-Riemann equations, 158
Cauchy-Schwarz inequality, 127, 128, 132, 181, 217, 244, 246
center of a group, 125, 126, 231
center of a pair of vectors, 118, 119
chain rule, 144, 153, 154
change of variables, 224, 226, 227, 230, 232
character, 246, 265
character group, 246, 250
characters, 240
Chevalley, Claude, 313
choice function, 306
classical compact groups, 125, 126
classical groups, 124
closed ball, 20, 62
closed convex hull, 43
Closed Graph Theorem, 110, 111
closed map, 96
closed rectangle, 181, 184, 185
closed set, 16, 21, 39, 41, 57, 65
closed subgroup, 98, 99, 123, 124
closed subspace, 105, 233
closed unit ball, 40
closure, 41, 42, 65, 98, 228, 267
Cobb-Douglas production function, 173
coefficients, 322
commutative algebra, 323
commutative ring, 289
commutative ring with identity, 87, 288
compact abelian group, 238, 251
compact group, 232, 246, 248
compact metric space, 77
compact open topology, 247, 248, 260
compact operator, 136, 138–140
compact set, 17, 56, 57, 59, 70
compact topological group, 98
compactly supported function, 103, 189, 220, 248, 255, 261
comparison test, 27
complement, 39, 275
complete independence of axioms, 310
complete measure, 200
complete metric space, 45, 47, 78, 103
complete ordered field, 72, 76
complete reducibility, 234
completion of a metric space, 63, 86
complex conjugate, 19
complex differentiable, 158
complex numbers, 19
complex power series, 28, 158, 163
component function, 154, 177
composite number, 285
composition of functions, 144, 292
congruence, 285
congruence modulo n, 308
conjugate, 94
conjugate exponent, 217
conjugate linear, 127
connected component, 62
connected set, 61, 159, 160
connected topological space, 97
consistent, 310
constant function, 291
constraint, 170, 171, 173–175
continuity of measure, 194
continuous function, 47, 48, 57, 61, 66, 70, 205
continuously differentiable, 147, 241
contraction mapping, 78, 79
Contraction Mapping Theorem, 78, 79
convergent infinite series, 133, 240
convergent sequence, 10, 13, 20, 44
convergent series, 26, 89
convex hull, 43
convex set, 43, 59, 61, 128, 163, 164, 177
convolution, 239, 248, 253, 256
coset, 89, 97
countable, 301
countable Cartesian product, 303
countably additive, 193, 198
countably infinite, 303
countably subadditive, 195, 196
counting measure, 193, 200, 216, 249
critical point, 167–171
cross product, 336

cup, 55
cyclic group, 341

decimal expansion, 29
Dedekind cut, 24
degree of a polynomial, 296
degree of a representation, 233
delta measure, 193, 200
DeMorgan's laws, 40, 276, 277
dense, 133
dense set, 5, 60, 63, 66, 81, 89, 123, 220, 245
derivative, 142, 146, 147, 149–154, 157, 158, 160–162, 167, 176
diagonal matrices, 344
diameter, 43, 118, 187
Dieudonné, Jean, 191
difference of sets, 274
differentiable, 142, 149, 150, 168, 191, 192
dihedral group, 346
dimension of a vector space, 318
Dini's theorem, 58, 83
direct sum, 234
directional derivative, 150, 151, 155, 156, 163
Dirichlet kernel, 242, 264
Dirichlet's theorem, 7
disconnected set, 61
discontinuity, 243, 264, 265
discontinuous, 243
discrete metric, 35, 38, 40–42, 48, 56, 62, 80, 200
discrete topology, 65, 74, 102, 248
disjoint sets, 274
distance between sets, 58, 177
divergent series, 26, 76
divisibility, 285
divisor, 285
domain of a function, 292
Dominated Convergence Theorem, 212–214, 216
dot product, 334
dual basis, 121
dual group, 247, 250, 251, 259, 261
dual space, 108, 109, 114, 117, 130

Egorov's theorem, 207, 208, 215
eigenspace, 137, 140
eigenvalue, 137
eigenvector, 137, 140
element of a set, 272
empty set, 272
endpoints of a rectangle, 184, 185
entire function, 255
epimorphism, 343
equality of sets, 272
equicontinuous, 93
equivalence class, 286
equivalence relations, 284
essential supremum, 217
essentially bounded, 217
Euler's theorem, 30
even permutation, 326
eventually constant sequence, 45, 86
Exchange Lemma, 317

face of a rectangle, 185
factorial, 284
Fatou's lemma, 211, 213, 215, 219
field, 288
field of rational functions, 73
finite complement topology, 65
finite field, 19, 89
finite rank operator, 138
finite set, 298, 302
finite-dimensional representation, 233
finite-dimensional vector space, 317
first category, 80
fixed point, 78
Fourier coefficient, 239
Fourier inversion, 249, 251, 252, 254, 256
Fourier series, 238, 239, 263
Fourier transform, 238, 242, 248, 251, 252, 261
Fréchet, Maurice René, 271
Fubini's theorem, 204, 220–224, 239, 240, 253
function equality, 292
fundamental counting principle, 278
Fundamental Theorem of Algebra, 71, 72
Fundamental Theorem of Calculus, 145

general linear group, 341, 344
generalized Cantor set, 80
generalized limit, 114, 116, 229
Generalized Mean Value Theorem, 145
generalized rotation, 347
geometric series, 26
Gibbs phenomenon, 264, 265
glb, 306
global maximum, 166
global minimum, 166
gradient vector, 155, 164, 171

Gram-Schmidt orthogonalization, 133, 339
greatest integer function, 295
greatest lower bound, 3, 306
greatest lower bound property, 3
group, 326, 340
group of norms, 95
group of units, 87

Haar basis, 258
Haar integral, 231
Haar measure, 194, 227, 240, 246
half-open rectangle, 181, 195, 203
Hamel basis, 319
harmonic series, 27
Hausdorff topological space, 67, 70, 96, 97
heat equation, 238
Heine-Borel theorem, 16, 22, 56, 57, 59, 88
Hessian, 168, 169
Hilbert space, 126, 128, 129, 131, 132, 134, 136, 140, 233, 234, 258, 267
Hilbert space adjoint, 130
Hölder's inequality, 35, 103, 217, 218, 222
homeomorphic, 77, 248
homeomorphism, 49, 57, 66, 70, 96, 201
homogeneous space, 96–98
homomorphism, 231, 233, 239, 246, 343
homomorphism of topological groups, 98, 99
hyperplane, 333

ideal in a ring, 87
idempotent operator, 129
identity, function, 291, 293
ILC, 288
image, 292
imaginary part, 19
Implicit Function Theorem, 175, 176, 179, 180
inclusion-exclusion principle, 279
independence of axioms, 310
index set, 277
indiscrete topology, 65
infimum, 55, 306
infinite series, 26, 89, 90
infinite set, 298, 302
infinite-dimensional representation, 233, 267
infinite-dimensional vector space, 319
infinitely differentiable function, 248
injective, 292
inner product, 126, 128, 129, 132, 257
inner regular Borel measure, 200
integrable function, 212–215, 220, 230
integral, 208, 209, 212, 213, 215, 251
integral domain, 288
interior, 42, 65, 80, 143, 181
Intermediate Value Theorem, 61, 176
Intermediate Value Theorem for Derivatives, 146
internal law of composition, 288
intersection, 274
intertwining operator, 235
interval, 8, 61
invariant subspace, 233, 235, 237
inverse Fourier transform, 251
Inverse Function Theorem, 175, 176, 178, 179
inverse of a function, 294
invertible matrix, 324
irrational number, 4, 33, 101
irreducible representation, 232, 233, 235, 265, 267, 269
isolated point, 40, 80
isometric, 49, 77
isometric isomorphism, 119
isometrically isomorphic, 258
isometry, 49, 116, 118, 119, 141, 238, 258
isometry group, 120, 121, 123
isomorphic, 248
isomorphism of groups, 343
isomorphism of topological groups, 98, 99
Iwasawa decomposition, 141

Jacobian matrix, 155, 170, 175, 224, 225
Jordan content zero, 189
Jordan domain, 189

kernel, 99, 108, 250
kernel of a homomorphism, 343
kernel of a linear transformation, 321

L'Hôpital's rule, 146, 147
Lagrange multiplier, 170–173
Laplace transform, 238, 255
lattice, 83, 84, 206
lattice operations, 55
lattice point, 1, 80
lattice polygon, 1
lattice polyhedron, 1
least upper bound, 2, 72, 306

least upper bound property, 12, 25, 42, 72, 74, 76
Lebesgue inner measure, 195, 226
Lebesgue integrable, 213
Lebesgue integral, 208, 212, 224
Lebesgue measurable set, 199, 202
Lebesgue measure, 192, 194, 200, 204, 229, 249
Lebesgue measure zero, 199–202, 205
Lebesgue outer measure, 194, 195, 226
Lebesgue's theorem, 182, 184
left Haar measure, 227–232
left regular representation, 235
left-hand derivative, 241
limit, 10, 13, 20, 44
limit infimum, 54, 55
limit supremum, 54, 55
line, 333
linear combination, 314
linear functional, 112
linear independence, 315
linear isometry, 117
linear isomorphism, 320
linear operator, 319
linear transformation, 106, 149, 150, 202, 319
linearly dependent, 316
linearly independent set, 132, 133, 179
Liouville's theorem, 148
Lipschitz condition, 79, 202
Little Hahn-Banach Theorem, 112
local maximum, 145, 148, 166–170, 172, 173
local minimum, 145, 148, 166–170, 172, 173
locally compact, 89, 123
locally compact abelian group, 232, 246–248, 265
locally compact group, 123, 124, 126, 194, 265
locally compact space, 70
locally compact topological group, 227
locally compact topological space, 99
locally constant function, 248, 261
lower bound, 2, 306
lub, 306
Lusin's theorem, 208

mathematical induction, 283
matrix, 121, 123, 154, 155, 162, 169, 179, 230, 268, 322
maximal compact subgroup, 126
maximal element, 306
maximal ideal, 87, 89, 95
maximum value, 144, 166, 174
Mazur-Ulam theorem, 118, 119, 203
Mean Value Theorem, 145, 148, 149, 156–161, 163, 177, 180
measurable, 196
measurable function, 192, 205, 206, 208, 209, 212
measurable set, 196
measure, 192, 193, 260
measure space, 193, 205, 208
measure zero, 182, 185, 187–189, 207
metric, 34, 86, 228
metric induced by the norm, 103
metric space, 34, 65, 199, 205
metric topology, 228
Michelle function, 293
minimum value, 144, 166, 174
Minkowski's inequality, 219
Minkowski's integral inequality, 221–223
mixed partial derivative, 160, 161
model, 310
monomorphism, 343
Monotone Convergence Theorem, 210, 211, 221, 222
monotonic function, 176, 182
monotonic increasing function, 52, 182
monotonic increasing sequence, 14
monotonic sequence of functions, 58
monotonicity of Lebesgue outer measure, 195
monotonicity of measure, 194
monotonicity of the integral, 209
motion group, 266
mutlinomial coefficient, 162

nearly continuous function, 182
negative definite quadratic form, 168
neighborhood, 9, 65, 96, 97
Nested Intervals Theorem, 9
non-Archimedean absolute value, 94
non-Archimedean triangle inequality, 85, 87, 92
nonmeasurable set, 204
nonnegative extended real numbers, 193
nonoverlapping intervals, 51
nonsingular linear transformation, 176–179, 202
norm, 102, 127, 334
normal subgroup, 98, 126, 231, 344
normal topological space, 67, 68
normalized vector, 337

normed linear space, 64, 102, 103, 116, 119, 219
nowhere dense set, 80

one-to-one, 292
one-to-one correspondence, 292
onto, 292
open ball, 20, 36
open cover, 17, 56
open interval, 73
open map, 66, 96, 99, 106
open mapping, 112
Open Mapping Theorem, 110, 111
open rectangle, 181, 188, 194
open set, 16, 21, 38, 48, 65, 198, 200
order axioms, 281
order isomorphic, 3
order of a group, 342
order of an element, 342
order topology, 73
ordered field, 2, 3, 42, 73, 86, 289
ordered integral domain, 2, 72, 73, 289
ordered pair, 277
ordinary least squares approximation, 168
orthogonal, 335, 337
orthogonal complement, 129, 140
orthogonal group, 123, 124, 344
orthogonal matrix, 204
orthogonal projection, 235
orthogonal set, 132
orthogonal transformation, 347
orthonormal basis, 132, 133, 140, 245, 258, 337
orthonormal set, 132, 133, 245
Osgood, William Fogg, 1
Ostrowski's theorem, 102
Other Hahn-Banach Theorem, 113–115, 117
outer measure, 196, 197, 228, 229
outer regular Borel measure, 200

pairwise disjoint, 274
paraboloid, 165
parallelepiped, 333
parallelogram law, 126, 128
Parseval's identity, 245, 249
partial derivative, 150, 154, 161, 162
partial differential equations, 238
partial sum, 26
partially ordered set, 305
partition, 185–188
perfect set, 80
permutation, 325
perpendicular, 337
Picard's theorem, 79
piecewise continuous, 244
piecewise continuous function, 241, 243, 244
piecewise smooth function, 241, 243, 244, 264, 265
pigeonhole principle, 292
Plancherel's theorem, 248, 249, 256, 257
plane, 333
point, 332
pointwise convergent sequence, 46, 51, 244
pointwise evaluation, 109
polar form, 23
polynomial function, 82, 296
Pontryagin-van Kampen Duality Theorem, 248, 250
positive definite bilinear form, 334
positive definite Hermitian form, 126–128
positive definite quadratic form, 168, 169
positive operator, 135, 136
power set, 279
preimage, 294
prime number, 285
primitive root of unity, 23, 92
principal branch of the argument, 23
principal units, 91
product measure, 224, 232
product rule, 152
product topology, 69, 70
projection, 69, 336
projection onto a closed subspace, 129, 131
projection onto the k-th coordinate, 151
proper subset, 272
pure imaginary, 19

quadratic form, 168, 169
quantifiers, 273
quotient field, 73
quotient map, 98, 99
quotient topology, 70, 266

radius of convergence, 29, 158, 163
ramified extension, 95
range, 292
rapidly decreasing at infinity, 248, 255
ratio test, 28
rational number, 101, 286

real algebraic number, 24, 72, 148, 149, 192
real numbers, 3, 86, 101, 142
real part, 19
rectangle, 181, 184, 194, 203
reduced residue system, 309
refinement, 186
reflexive Banach space, 117
regular Borel measure, 200
regular topological space, 67, 68, 97
Reisz-Fisher theorem, 246
relation, 278
relative field norm, 94
relative topology, 42, 69
remainder term, 148, 163
representation, 233, 265, 267
residue field, 95
resolvent set, 137
Riemann integrable, 209, 212
Riemann integrable function, 186–188, 209
Riemann integral, 186, 188, 209, 212
Riemann's localization principle, 255
Riemann-Lebesgue lemma, 239, 253
Riesz Representation Theorem, 130
Riesz's lemma, 139, 140
Riesz-Fisher theorem, 245
right Haar integral, 231
right Haar measure, 227, 228, 230–232
right regular representation, 235, 237
right-hand derivative, 241
ring of integers, 87–89, 95
Rolle's theorem, 145, 148, 176
root test, 28
rules of arithmetic, 280

saddle point, 169, 170
sawtooth curve, 264
sawtooth function, 265
scalar, 314
scalar multiplication, 314
scalar product, 334
Schröder-Bernstein theorem, 299
Schwartz space, 248, 255, 261, 267
second category, 80, 81
second derivative test, 168, 169
self-dual, 250
selfadjoint operator, 131, 136, 140
semidirect product, 346
separable metric space, 60, 133, 134
separable topological space, 67, 227
separates points, 82
sequence, 297
sequential covering class, 228
sequentially compact set, 17, 23, 58, 59
set containment, 272
shell, 87, 89, 90
sign of a permutation, 326
simple function, 207, 261
smooth function, 162, 248
smooth hypersurface, 165, 166, 172
space of left cosets, 98, 99
span, 316
spanning set, 316
special linear group, 124, 343, 344
special orthogonal group, 124, 347
special unitary group, 125
Spectral Theorem for Compact Selfadjoint Operators, 136, 140
spectrum, 137, 138
standard basis, 121, 132, 150, 315, 317
step function, 253
Stone's theorem, 83, 84
Stone-Weierstrass theorem, 78, 82, 85, 104
strictly monotonic decreasing function, 145
strictly monotonic increasing function, 52, 145
strongly continuous representation, 233
strongly convergent sequence, 134
subadditive, 196
subadditive functional, 113–116
subbasis for a topology, 248
subfield, 18, 73, 289
subgroup, 98, 342
subintegral domain, 289
subrepresentation, 233
subring, 289
subset, 272
subspace, 318
subspace topology, 69
superposition, 238
superset, 272
support, 103
supremum, 3, 55, 208, 306
surjective, 292
symmetric bilinear form, 334
symmetric difference, 275
symmetric group, 326
symmetric neighborhood, 9
symplectic group, 124

tangent hyperplane, 165
tangent line, 149, 164
taxicab metric, 34

Taylor expansion, 148
Taylor polynomial, 163, 164, 169, 180
Taylor's theorem, 147, 148, 163
terminating decimal expansion, 30
Toeplitz, Otto, 141
topological basis, 104, 132
topological group, 96, 97, 99, 123
topological space, 48, 64, 96
topologically isomorphic, 248, 250, 251, 259
topology, 39, 64, 228, 232
topology generated by a collection of sets, 66
total variation, 52
totally bounded, 58
totally disconnected, 62
totally ordered set, 305
transcendental number, 24, 192
translation, 117, 203
transpose, 121
triangle inequality, 10, 13, 20, 34, 102
trigonometric polynomial, 85, 245
trigonometric series, 238
Tychonoff's theorem, 70

uniform boundedness principle, 81, 109
uniformly absolutely continuous, 215
uniformly bounded sequence of functions, 213
uniformly continuous function, 51, 57
uniformly convergent sequence, 46, 50, 58, 82, 134, 244
unimodular group, 231, 232, 235, 265
union, 273
unipotent matrix, 204, 225
unit ball, 37, 111
unit circle, 22, 238
unit group, 95
unit interval, 304
unit sphere, 165
unit vector, 337
unitarily equivalent, 235, 267
unitary character, 239, 246, 247, 265
unitary dual, 234, 267
unitary equivalence, 234, 270
unitary group, 125
unitary operator, 131, 233, 238
unitary representation, 232, 233, 267
units in a ring, 82, 87
universal set, 273
unramified extension, 95
upper bound, 2, 306
upper triangular matrices, 344
upper triangular unipotent matrices, 344
Urysohn's lemma, 68, 189
Urysohn's Metrization Theorem, 68
usual metric on $\mathbb{R}^n$, 34

valuation, 89, 90
variation, 52
vector space, 256, 313
Vitali Convergence Theorem, 215
volume of a rectangle, 194, 197

wave equation, 238
weakly continuous representation, 233
weakly convergent sequence, 134
Weierstrass's theorem, 82
Weil, André, 77
well-defined, 287
well-ordered set, 305
well-ordering principle for $\mathbb{Z}$, 283

Zahlen, 280
Zygmund, Antoni, 237

Index of Notation Definitions

$[\cdot]$, 295
$\bigvee$, 55
$\bigwedge$, 55
∂A, 40
$\frac{\partial f}{\partial x_j}$, 151
$\int_R f(x)dx$, 187
$\int_X f d\mu$, 208
$\nabla f(x)$, 155
$\varnothing$, 272
$|\cdot|_\theta$, 94
$|\cdot|_p$, 85
$\|\cdot\|$, 34, 107
$\|\cdot\|_\infty$, 36, 103
$\|\cdot\|_p$, 35, 102
$\|\cdot\|_p$, $1 \leq p \leq \infty$, 217
A°, 42
A_n, 326
$\mathbb{A}$, 24
$\mathbb{A}_\mathbb{R}$, 24
$\mathrm{Aut}\,(F)$, 18
$\overline{A}$, 41
${}^tA$, 327
B^n, 40
$B_r(x)$, 36
$\overline{B}_r(x_0)$, 36
$\mathcal{BC}(X, F)$, 50
$\mathcal{BL}(V)$, 108
$\mathcal{BL}(V, W)$, 107, 108
$\mathcal{B}(X, F)$, 44
$C^\infty(a, b)$, 147
$C^k(U)$, 147, 162
$C^k(a, b)$, 147
$\mathbb{C}$, 19
$\mathbf{c}$, 303
$\mathcal{C}(X, F)$, 103
$\mathcal{C}_c(X, F)$, 103
$c(F)$, 104
$c_0(F)$, 104
D^+f, 183
D^-f, 183
$D_N(\theta)$, 242
$D_j f(x)$, 150
$D_v f(x)$, 150
$D_{ij} f(x)$, 161
$Df(x)$, 142, 150
$\det A$, 327
$\mathrm{diam}(A)$, 43
$d(A, B)$, 58
$d(x, y)$, 34
d^+f, 183
d^-f, 183
$d_\infty(x, y)$, 36
$d_p(x, y)$, 35
e_j, 104
$\mathbb{F}_\theta$, 95
$\mathbb{F}_p$, 19
$\hat{f}$, 239
$\mathscr{F}$, 252
$\mathcal{F}(X, F)$, 315
$\tilde{f}$, 240
$f'(x)$, 142
$f * g$, 222, 239
f^+, 182, 206
f^-, 182, 206
$GL_n(F)$, 344
glb, 3
$\hat{G}$, 247
$H_x(v)$, 168
inf, 3
$\mathcal{I}(V)$, 120
$J(x)$, 224
ker, 321, 343
$L(f, P)$, 186
$L^p(\mu)$, 217
L. i. m., 114
$\ell^p, 1 \leq p \leq \infty$, 102
$\ell_n^p, 1 \leq p \leq \infty$, 36
lim inf, 54
lim sup, 54
lub, 3
$\mathscr{L}(f)$, 255
$\mathscr{L}(f, R)$, 186

$M_f(R)$, 186
m^*, 195
m_*, 195
$m_f(R)$, 186
$\mathbb{N}$, 279
$\mathrm{Nm}_{\mathbb{Q}_p(\sqrt{\theta})/\mathbb{Q}_p}(z)$, 94
$O_n(F)$, 344
$\mathfrak{p}$, 87
$\mathfrak{p}_\theta$, 95
$\mathrm{proj}_{\mathbf{w}}(\mathbf{v})$, 336
$\mathscr{P}(X)$, 279
p_k, 151
$\mathbb{Q}$, 286
$\mathbb{Q}_p$, 86, 94, 101
$\mathbb{Q}_{p,\theta}^\times$, 95
$R(x)$, 73
R_θ, 95
R_p, 87
$\mathbb{R}$, 3
$\mathbf{R}$, 11
$SL_n(F)$, 344
$SO_n(\mathbb{R})$, 347
S^n, 40
$S_N(f)$, 240
$S_R(x)$, 254
S_n, 325
$\mathcal{S}(\mathbb{R})$, 255
$\mathcal{S}_n$, 267
$\mathrm{sgn}(\sigma)$, 326
sup, 3
$\mathrm{supp} f$, 103
T^*, 116
T_χ, 266
$\mathbb{T}$, 238
$U(f, P)$, 186
U_θ, 95
U_p, 87
$U_{p,n}$, 90
$\mathscr{U}(f, R)$, 187
V^*, 108
$V^\perp$, 129
$\mathrm{var}(f, P)$, 53
$\mathbb{Z}$, 280
$\mathbb{Z}_n$, 290, 340
$\aleph_0$, 298
Δ, 231
$\lambda(x)$, 90
μ^*, 196
$\nu(x)$, 89
ϕ, 309
φ_n, 258
$\rho(T)$, 137
$\sigma(T)$, 137
χ_m, 251
χ_n, 238
χ_t, 251
$\psi(x)$, 258, 259
$\psi_u(x)$, 259
$\psi_{j,k}(x)$, 258

This is a textbook for a course in Honors Analysis (for freshman/sophomore undergraduates) or Real Analysis (for junior/senior undergraduates) or Analysis-I (beginning graduates). It is intended for students who completed a course in "AP Calculus", possibly followed by a routine course in multivariable calculus and a computational course in linear algebra.

There are three features that distinguish this book from many other books of a similar nature and which are important for the use of this book as a text. The first, and most important, feature is the collection of exercises. These are spread throughout the chapters and should be regarded as an essential component of the student's learning. Some of these exercises comprise a routine follow-up to the material, while others challenge the student's understanding more deeply. The second feature is the set of independent projects presented at the end of each chapter. These projects supplement the content studied in their respective chapters. They can be used to expand the student's knowledge and understanding or as an opportunity to conduct a seminar in Inquiry Based Learning in which the students present the material to their class. The third really important feature is a series of challenge problems that increase in impossibility as the chapters progress.

ISBN 978-0-8218-9141-4
9 780821 891414
AMSTEXT/20

For additional information and updates on this book, visit
www.ams.org/bookpages/amstext-20

AMS on the Web
www.ams.org

The Sally SERIES

This series was founded by the highly respected mathematician and educator, Paul J. Sally, Jr.